

PCTWORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁷ : A61K 39/395, C12P 7/62, C12N 9/02, 15/00, C07H 19/00	A1	(11) International Publication Number: WO 00/21557 (43) International Publication Date: 20 April 2000 (20.04.00)
(21) International Application Number: PCT/US99/23253 (22) International Filing Date: 5 October 1999 (05.10.99) (30) Priority Data: 60/103,760 9 October 1998 (09.10.98) US (71) Applicant (for all designated States except US): MERCK & CO., INC. [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065 (US). (72) Inventors; and (75) Inventors/Applicants (for US only): PETRUKHIN, Konstantin [RU/US]; 126 East Lincoln Avenue, Rahway, NJ 07065 (US). CASKEY, C., Thomas [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065 (US). (74) Common Representative: MERCK & CO., INC.; 126 East Lincoln Avenue, Rahway, NJ 07065 (US).		(81) Designated States: CA, JP, US, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). Published <i>With international search report.</i> <i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>
(54) Title: DELTA 6 FATTY ACID DESATURASE (57) Abstract Novel human DNA sequences that encode the gene CYB5RP, a delta 6 fatty acid desaturase, are provided. Provided are genomic CYB5RP DNA as well as cDNA that encodes the CYB5RP protein. Also provided is CYB5RP protein encoded by the novel DNA sequences. Methods of expressing CYB5RP protein in recombinant systems are provided. Also provided are CYB5RP methods that identify activators and inhibitors of CYB5RP protein.		

BEST AVAILABLE COPY

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

TITLE OF THE INVENTION
DELTA 6 FATTY ACID DESATURASE

CROSS-REFERENCE TO RELATED APPLICATIONS

5 Not applicable.

STATEMENT REGARDING FEDERALLY-SPONSORED R&D

Not applicable.

10 REFERENCE TO MICROFICHE APPENDIX

Not applicable.

FIELD OF THE INVENTION

15 The present invention is directed to novel human DNA sequences encoding a delta 6 fatty acid desaturase, an enzyme involved in the synthesis of essential fatty acids.

BACKGROUND OF THE INVENTION

20 Essential fatty acids (EFAs) are polyunsaturated fatty acids that cannot be manufactured by mammals, yet are required for a number of important biochemical processes, and thus must be supplied in the diet. The most important dietary EFAs are linoleic acid and alpha-linolenic acid (ALA). These two EFAs undergo a number of biosynthetic reactions that convert them into various other EFAs. Figure 1 depicts the biosynthetic reactions involving the two groups of EFAs, the n-6 EFAs (linoleic acid derivatives) and the n-3 EFAs (ALA derivatives). EFAs are formed from linoleic acid and ALA by a series of alternating reactions involving the removal of two
25 hydrogens coupled with the insertion of an additional double bond (desaturation) and the lengthening of the fatty acid chain by the addition of two carbons (chain elongation). The enzymes catalyzing the desaturations and elongations are thought to
30 be the same for both groups of EFAs.

Among the more important unsaturated fatty acids are the delta 6 unsaturated fatty acids, which are involved in the maintenance of membrane structure and function, the regulation of cholesterol synthesis and transport, and the prevention

of water loss from the skin. Delta 6 unsaturated fatty acids also serve as precursors of the eicosanoids, including the prostaglandins and leukotrienes (Horrobin, 1992, Prog. Lipid Res. 31:163-194). The double bond at the 6 position of delta 6 unsaturated fatty acids is introduced by a class of enzymes known as delta 6 desaturases.

5 Deficiencies in linoleic acid and ALA derivatives have been associated with skin diseases, diabetic complications, inflammatory and autoimmune disorders, cardiovascular disorders, complications of viral infection, and retinal dysfunction. For example, a deficiency in gamma-linolenic acid (GLA), which is produced from linoleic acid by the action of the enzyme delta 6 desaturase, can arise from the
10 decreased activity of this enzyme that occurs in aging, stress, diabetes, eczema, and some infections, or from increased catabolism of GLA due to oxidation or rapid cell division, as occurs in inflammation or cancer. Clinical trials have demonstrated that dietary GLA supplementation can be effective in treating a number of conditions that are associated with GLA deficiency, *e.g.*, atopic eczema, mastalgia, diabetic
15 neuropathy, viral infections, and some forms of cancer (Horrobin, 1990, Rev. Contemp. Pharmacother. 1:1-45).

Delta 6 desaturase is an example of a fatty acid desaturase. Fatty acid desaturases are enzymes that introduce a double bond into the carbon chain of fatty acids. They play vital roles in the biosynthesis of polyunsaturated fatty acids,
20 including the essential fatty acids. Fatty acid desaturases are present in soluble and membrane-associated forms and require electron donors (for example, cytochrome b5) for their functioning.

Delta 6 desaturases catalyze the rate-limiting steps in the biosyntheses of the linoleic and ALA group EFAs shown in Figure 1. End products of the linoleic
25 acid pathway include the eicosanoids (prostaglandins and leukotrienes). The end product of the ALA pathway is docosahexaenoic acid (DHA), an important component of membranes in the vertebrate retina. DHA is highly specific for retina and represents more than 50% of the fatty acids in the rod outer segment (ROS). It appears that DHA is important in maintaining the normal structure and function of the
30 retina (Anderson et al., 1992, Neurobiology of Essential Fatty Acids, Bazan et al., eds., Plenum Press, New York, pages 285-294). Increased dietary consumption of DHA and its precursor, eicosapentaenoic acid, from seal meat and fish has been

linked to an increased incidence of macular degeneration in Greenland Eskimos (Rosenberg, 1987, *Arct. Med. Res.* 46:64-70).

Certain delta 6 desaturases have been cloned from plants. For example, a delta 6 desaturase has been cloned from borage (Sayanova et al., 1997, *Proc. Natl. Acad. Sci. USA* 94:4211-4216). This delta 6 desaturase is unusual in that its cytochrome b5 electron donor is present as an N-terminal extension of the enzyme rather than being synthesized as a separate protein. The borage delta 6 desaturase has been shown to be functional, in that transfer of the cloned gene encoding it to tobacco results in the synthesis of high levels of GLA and octadecatetraenoic acid (OTA) in the transgenic tobacco leaves. GLA and OTA are the products of delta 6 desaturase activity on linoleic acid and ALA, respectively.

Based on its hydropathy profile, the borage delta 6 desaturase appears to be a membrane-bound protein. Examination of the amino acid sequence of the borage enzyme, as well as the amino acid sequences of membrane-bound desaturases from a wide variety of organisms, has revealed three regions of conserved short motifs containing histidine residues (HX(3 or 4)H, HX(2 or 3)HH, and HX(2 or 3)HH) having a conserved spacing from each other (Shanklin et al., *Biochemistry*, 1994, 33:12787-12794).

A DNA sequence has been isolated from sunflower embryos that, judging from its sequence, appears to encode a delta 6 desaturase having a cytochrome b5-like moiety fused to its N-terminus (Sperling et al., 1995, *Eur. J. Biochem.* 232:798-805).

SUMMARY OF THE INVENTION

The present invention is directed to novel human DNA sequences that encode a delta 6 fatty acid desaturase, cytochrome b5-related protein (CYB5RP). The present invention includes genomic CYB5RP DNA as well as cDNA that encodes the CYB5RP protein. The genomic CYB5RP DNA is substantially free from other nucleic acids and has the nucleotide sequence shown in SEQ.ID.NO.:1. The cDNA encoding CYB5RP protein is substantially free from other nucleic acids and has the nucleotide sequence shown in SEQ.ID.NO.:2. Also provided is CYB5RP protein encoded by the novel DNA sequences. The CYB5RP protein is substantially free from other proteins and has the amino acid sequence shown in SEQ.ID.NO.:3.

Methods of expressing CYB5RP protein in recombinant systems are provided. Also provided are methods of producing delta 6 unsaturated fatty acids using DNA encoding CYB5RP or using CYB5RP protein.

5 BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 depicts the enzymatic conversions involved in the linoleic acid (n-3) and alpha-linolenic acid (n-6) pathways of essential fatty acid synthesis.

Figure 2A-G shows the genomic DNA sequence of the CYB5RP gene (SEQ.ID.NO.:1). Underlined nucleotides in capitals represent exons. The start ATG codon at position 544 in exon 1 and the stop TGA codon at position 18,103 in exon 12 are shown in bold. The putative polyadenylation signal ATTAAA located approximately 20 base pairs upstream of the polyA tail is shown in bold italics (position 18,373 in exon 12). DNA sequence upstream of exon 1 represents a putative promoter region of the CYB5RP gene., as indicated by the presence of the TATA box at position 353 (underlined bold)..

Figure 3A-C shows the cDNA sequence (SEQ.ID.NO.:2) and the amino acid sequence (SEQ.ID.NO.:3) of CYB5RP. The region encompassing amino acids 1-102 represents the cytochrome b5 domain. The region encompassing amino acids 182-186 represents HIS BOX 1. The region encompassing amino acids 219-223 represents HIS BOX 2. The region encompassing amino acids 383-387 represents HIS BOX 3.

Figure 4 shows a portion of the cDNA sequence (SEQ.ID.NO.:4) and a portion of the amino acid sequence (SEQ.ID.NO.:5) of mouse CYB5RP.

Figure 5A shows a Kyte-Doolittle hydropathy plot of CYB5RP. Figure 5B shows the proposed membrane topology of CYB5RP based on its hydropathy plot. This membrane topology is similar to that proposed for other membrane-bound fatty acid desaturases (Shanklin et al., Biochemistry, 1994, 33:12787-12794). The amino acids shown in Figure 5B are portions of (SEQ.ID.NO.:3).

Figure 6 shows the output of the Profilescan program from the Wisconsin GCG package. The upper amino acid sequence is from CYB5RP (positions 31-78 of SEQ. ID. NO.3). The lower amino acid sequence is positions 1-48 of the cytochrome b5 profile (SEQ. ID. NO.:6.). The output shows that CYB5RP

contains a profile typical for the heme-binding domain of the cytochrome b5 protein family. Importantly, the region of identity includes the invariant HPGG motif, where histidine represents a heme axial ligand for iron.

Figure 7A and B show the results of BlastP searches of the GenBank database using the full-length CYB5RP amino acid sequence as the query. Figure 7A shows the hit with highest homology, a hypothetical protein from sunflower. The sunflower protein and CYB5RP share three His boxes (boxed) in which the spacing between the His boxes is conserved. Also boxed is the HPGG motif typical for the heme-binding domain of the cytochrome b5 protein family. In both proteins the first histidine of the third His box is replaced by glutamine (a typical feature of desaturases with delta 6 specificity). The upper amino acid sequences shown are from CYB5RP and are portions of SEQ. ID. NO.3. The lower amino acid sequences shown are portions of the amino acid sequence of the hypothetical protein from sunflower (Sperling et al., 1995, Eur. J. Biochem. 232:798-805). The sequence shown as positions 348-432 is SEQ. ID. NO.:7. The sequence shown as positions 22-74 is SEQ. ID. NO.:8. The sequence shown as positions 152-227 is SEQ. ID. NO.:9. Figure 7B shows the hit with the second highest homology, a delta 6 desaturase from *Borago officinalis* (Sayanova et al., 1997, Proc. Natl. Acad. Sci. USA 94:4211-4216). The *Borago* protein and CYB5RP also share three His boxes with conserved spacing, as well as the HPGG motif. In both proteins the first histidine of the third His box is replaced by glutamine (a typical feature of desaturases with delta 6 specificity). The upper amino acid sequences shown are from CYB5RP and are portions of SEQ. ID. NO.3. The lower amino acid sequences shown are portions of the amino acid sequence of the *Borago* delta 6 desaturase. The sequence shown as positions 338-424 is SEQ. ID. NO.:10. The sequence shown as positions 12-64 is SEQ. ID. NO.:11. The sequence shown as positions 153-220 is SEQ. ID. NO.:12.

Figure 8 shows additional results of BlastP searches of the GenBank database using the CYB5RP protein as the query. Figure 8 shows the amino acid alignment between the CYB5RP protein and a delta 6 desaturase from *Synechocystis* sp. (strain pcc 6803) performed by the BlastP program. The *Synechocystis* delta 6 desaturase and CYB5RP share three His boxes, two of which are shown in Figure 8 (boxed). In both proteins the first histidine of the third His box is replaced by glutamine (a typical feature of desaturases with delta 6 specificity). The CYB5RP

sequence shown is a portion of SEQ. ID. NO.3. The *Synechocystis* sequence shown is SEQ. ID. NO:13.

Figure 9A shows the expression pattern of the CYB5RP gene in 9 human tissues, as determined by RT-PCR amplification with 21 cycles. Expression is detected in human retina, kidney, pancreas, placenta, and brain. Figure 9B shows the results of the analogous experiments performed with 25 cycles of amplification. Expression of the CYB5RP gene is seen in all the human tissues studied.

DETAILED DESCRIPTION OF THE INVENTION

For the purposes of this invention:

"Substantially free from other proteins" means at least 90%, preferably 95%, more preferably 99%, and even more preferably 99.9%, free of other proteins. Thus, a CYB5RP protein preparation that is substantially free from other proteins will contain, as a percent of its total protein, no more than 10%, preferably no more than 5%, more preferably no more than 1%, and even more preferably no more than 0.1%, of non-CYB5RP proteins. Whether a given CYB5RP protein preparation is substantially free from other proteins can be determined by such conventional techniques of assessing protein purity as, e.g., sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) combined with appropriate detection methods, e.g., silver staining or immunoblotting.

"Substantially free from other nucleic acids" means at least 90%, preferably 95%, more preferably 99%, and even more preferably 99.9%, free of other nucleic acids. Thus, a CYB5RP DNA preparation that is substantially free from other nucleic acids will contain, as a percent of its total nucleic acid, no more than 10%, preferably no more than 5%, more preferably no more than 1%, and even more preferably no more than 0.1%, of non-CYB5RP nucleic acids. Whether a given CYB5RP DNA preparation is substantially free from other nucleic acids can be determined by such conventional techniques of assessing nucleic acid purity as, e.g., agarose gel electrophoresis combined with appropriate staining methods, e.g., ethidium bromide staining, or by sequencing.

"Substantially the same biological activity as CYB5RP" means being able to introduce a double bond into the 6 position of linoleic acid under conditions in which CYB5RP is able to introduce a double bond into the 6 position of linoleic acid.

A "conservative amino acid substitution" refers to the replacement of one amino acid residue by another, chemically similar, amino acid residue. Examples of such conservative substitutions are: substitution of one hydrophobic residue (isoleucine, leucine, valine, or methionine) for another; substitution of one polar residue for another polar residue of the same charge (e.g., arginine for lysine; glutamic acid for aspartic acid); substitution of one aromatic amino acid (tryptophan, tyrosine, or phenylalanine) for another.

The present invention relates to the identification and cloning of cytochrome b5-related protein (CYB5RP), a gene which encodes a human delta 6 fatty acid desaturase. The gene is present on PAC clones 759J12, 756B3, 519O13, and 466A11 from an area of human chromosome 11q12 that has been shown to contain a gene related to Best's macular dystrophy (Cooper *et al.*, 1997, Genomics 41:185-192; Stöhr *et al.*, 1997, Genome Res. 8:48-56; Graff *et al.*, 1997, Hum. Genet. 101: 263-279). This linkage between the chromosomal location of the CYB5RP gene and the location of the gene related to Best's macular dystrophy can be used diagnostically by identifying restriction fragment length polymorphisms (RFLPs) in the vicinity of the CYB5RP gene, e.g., in SEQ.ID.NO.:1. Such RFLPs will be associated with the Best's macular dystrophy gene and thus can be used to identify individuals carrying disease-causing forms of the Best's macular dystrophy gene.

CYB5RP was identified as an EST hit in sequence scanning data from PAC clones from human chromosome 11q12. In addition, a full length cDNA of CYB5RP was recovered from a human retina cDNA library. The genomic region of CYB5RP has been sequenced and the exon/intron organization of CYB5RP has been determined. The CYB5RP gene has 12 exons. The promoter region of CYB5RP was identified upstream of the 5' UTR by detecting consensus elements required for eukaryotic transcription. The expression pattern of CYB5RP was determined by RT-PCR analysis in 9 human tissues. The CYB5RP gene is expressed predominantly in human retina, kidney, pancreas, and placenta; lower levels of expression are also detected in brain, heart, lung, liver, and skeletal muscle. Bioinformatic analysis revealed significant homology to a group of plant and bacterial fatty acid desaturases. All of the typical amino acid motifs present in these fatty acid desaturases are also present in CYB5RP. Kyte-Doolittle algorithm analysis predicts a transmembrane organization typical of fatty acid desaturases for CYB5RP (see Figure 5). CYB5RP is

unusual in that it contains a cytochrome b5 region in its N terminus. While many fatty acid desaturases utilize cytochrome b5 as an electron donor, most have not incorporated this cytochrome as part of their polypeptide chain.

That CYB5RP is a fatty acid desaturase is shown by the following
5 evidence:

- (1) CYB5RP possesses significant homology to a group of plant and microbial fatty acid desaturases;
- (2) Like other fatty acid desaturases, CYB5RP has three conserved histidine boxes, with correct spacing between the boxes; and
- 10 (3) The predicted membrane topology of CYB5RP is similar to that of known fatty acid desaturases.

That CYB5RP is a delta 6 fatty acid desaturase is shown by the following evidence:

- 15 (1) CYB5RP contains a cytochrome b5-like moiety fused to its N-terminus. The only two fatty acid desaturases that contain cytochrome b5-like moiety fused to their N-termini are known or suspected to be delta 6 desaturases.
- (2) The only two plant desaturases that are known or suspected to introduce a double bond in the 6 position have an atypical His box 3 (QI/LEHH), with a Q in the first position rather than an H. CYB5RP has the same atypical His Box 3.
- 20 (3) The only bacterial desaturase that is known to introduce a double bond in the 6 position has an atypical His box 3 (QVTHH), with a Q in the first position rather than an H. CYB5RP has the same atypical His Box 3.

CYB5RP is a target for the development of drugs for the treatment of disorders of lipid metabolism and for the treatment of conditions that require the
25 modulation of the biosynthesis of prostaglandins and leukotrienes (asthma, pain, *etc.*). CYB5RP is also a target for the development of drugs for use in treating skin diseases, diabetic complications, reproductive disorders, including breast pain and premenstrual syndrome, inflammatory and autoimmune disorders, cardiovascular disorders, complications of viral infections, and various forms of retinal degeneration,
30 including age-related macular degeneration.

CYB5RP is homologous to a delta 6 desaturase from *Borago officinalis* (see Figure 7B). Both CYB5RP and this *Borago* delta 6 desaturase, unlike desaturases from higher plants, are unusual in containing a cytochrome b5-like

domain fused to their N-termini (Sayanova et al., 1997, Proc. Natl. Acad. Sci. USA 94:4211-4216; hereinafter "Sayanova"). The *Borago* desaturase has been expressed in transgenic tobacco, resulting in high levels of delta 6 desaturated fatty acids in the transgenic tobacco leaves, including high levels of γ -linolenic acid (GLA) (Sayanova).

- 5 Given the medical importance of GLA, Sayanova proposed that transgenic plants, expressing the *Borago* delta 6 desaturase, would be valuable as sources of GLA. Similarly, CYB5RP, expressed in transgenic plants, is expected to provide a valuable source of GLA.

The present invention provides DNA encoding CYB5RP that is
10 substantially free from other nucleic acids. The present invention also provides recombinant DNA molecules encoding CYB5RP. The present invention provides DNA molecules substantially free from other nucleic acids comprising the nucleotide sequence shown in Figure 2 as SEQ.ID.NO.:1. Analysis of SEQ.ID.NO.:1 revealed that this genomic sequence defines a gene having 12 exons. These exons collectively
15 have an open reading frame that encodes a protein of 445 amino acids. When an alternatively spliced exon 8 is used, a CYB5RP protein of 433 amino acids, lacking amino acids 317-328, is produced. Thus, the present invention includes two cDNA molecules, encoding two forms of CYB5RP protein, that are substantially free from other nucleic acids. The first cDNA is shown in Figure 3 and has the nucleotide
20 sequence SEQ.ID.NO.:2. The second cDNA is identical to the first, except that it does not contain the nucleotides at positions 1,019-1,054.

The present invention includes DNA molecules substantially free from other nucleic acids comprising the coding region of SEQ.ID.NO.:2. Accordingly, the present invention includes DNA molecules substantially free from other nucleic acids
25 having a sequence comprising positions 71-1,405 of SEQ.ID.NO.:2. The present invention also includes DNA molecules substantially free from other nucleic acids having a sequence comprising positions 71-1,405 of SEQ.ID.NO.:2, except that the nucleotides at positions 1,019-1,054 are missing. Also included in the present invention are recombinant DNA molecules having a nucleotide sequence comprising
30 positions 71-1,405 of SEQ.ID.NO.:2 and recombinant DNA molecules having a nucleotide sequence comprising positions 71-1,405 of SEQ.ID.NO.:2 with the exception that positions 1,019-1,054 are missing.

The novel DNA sequences of the present invention encoding CYB5RP, in whole or in part, can be linked with other DNA sequences, *i.e.*, DNA sequences to which CYB5RP is not naturally linked, to form "recombinant DNA molecules" encoding CYB5RP. Such other sequences can include DNA sequences that control transcription or translation such as, *e.g.*, translation initiation sequences, promoters for RNA polymerase II, transcription or translation termination sequences, enhancer sequences, sequences that control replication in microorganisms, sequences that confer antibiotic resistance, or sequences that encode a polypeptide "tag" such as, *e.g.*, a polyhistidine tract or the myc epitope. The novel DNA sequences of the present invention can be inserted into vectors such as plasmids, cosmids, viral vectors, P1 artificial chromosomes, or yeast artificial chromosomes.

Included in the present invention are DNA sequences that hybridize to at least one of SEQ.ID.NOs.:1 or 2 under stringent conditions. By way of example, and not limitation, a procedure using conditions of high stringency is as follows:

Prehybridization of filters containing DNA is carried out for 2 hr. to overnight at 65°C in buffer composed of 6X SSC, 5X Denhardt's solution, and 100 µg/ml denatured salmon sperm DNA. Filters are hybridized for 12 to 48 hrs at 65°C in prehybridization mixture containing 100 µg/ml denatured salmon sperm DNA and 5-20 X 10⁶ cpm of ³²P-labeled probe. Washing of filters is done at 37°C for 1 hr in a solution containing 2X SSC, 0.1% SDS. This is followed by a wash in 0.1X SSC, 0.1% SDS at 50°C for 45 min. before autoradiography.

Other procedures using conditions of high stringency would include either a hybridization carried out in 5XSSC, 5X Denhardt's solution, 50% formamide at 42°C for 12 to 48 hours or a washing step carried out in 0.2X SSPE, 0.2% SDS at 65°C for 30 to 60 minutes.

Reagents mentioned in the foregoing procedures for carrying out high stringency hybridization are well known in the art. Details of the composition of these reagents can be found in, *e.g.*, Sambrook, Fritsch, and Maniatis, 1989, Molecular Cloning: A Laboratory Manual, second edition, Cold Spring Harbor Laboratory Press. In addition to the foregoing, other conditions of high stringency which may be used are well known in the art.

The degeneracy of the genetic code is such that, for all but two amino acids, more than a single codon encodes a particular amino acid. This allows for the

construction of synthetic DNA that encodes the CYB5RP protein where the nucleotide sequence of the synthetic DNA differs significantly from the nucleotide sequence of SEQ.ID.NO.:2, but still encodes the same CYB5RP protein shown as SEQ.ID.NO.:3. Such synthetic DNAs are intended to be within the scope of the present invention. Also with the scope of the present invention are synthetic DNAs that encode a CYB5RP protein lacking amino acids 317-328 of SEQ.ID.NO.:3.

Another aspect of the present invention includes host cells that have been engineered to contain and/or express DNA sequences encoding CYB5RP protein. Such recombinant host cells can be cultured under suitable conditions to produce CYB5RP protein. An expression vector containing DNA encoding CYB5RP protein can be used for expression of CYB5RP protein in a recombinant host cell. Recombinant host cells may be prokaryotic or eukaryotic, including but not limited to, bacteria such as *E. coli*, fungal cells such as yeast, mammalian cells including, but not limited to, cell lines of human, bovine, porcine, monkey and rodent origin, plant cells such as tobacco, and insect cells including but not limited to *Drosophila* and silkworm derived cell lines. Cell lines derived from mammalian species which are suitable for recombinant expression of CYB5RP protein and which are commercially available, include but are not limited to, L cells L-M(TK⁻) (ATCC CCL 1.3), L cells L-M (ATCC CCL 1.2), 293 (ATCC CRL 1573), Raji (ATCC CCL 86), CV-1 (ATCC CCL 70), COS-1 (ATCC CRL 1650), COS-7 (ATCC CRL 1651), CHO-K1 (ATCC CCL 61), 3T3 (ATCC CCL 92), NIH/3T3 (ATCC CRL 1658), HeLa (ATCC CCL 2), C1271 (ATCC CRL 1616), BS-C-1 (ATCC CCL 26) and MRC-5 (ATCC CCL 171).

A variety of mammalian expression vectors can be used to express recombinant CYB5RP in mammalian cells. Commercially available mammalian expression vectors which are suitable include, but are not limited to, pMC1neo (Stratagene), pSG5 (Stratagene), pcDNA1 and pcDNA1amp, pcDNA3, pcDNA3.1, pCR3.1 (Invitrogen), EBO-pSV2-neo (ATCC 37593), pBPV-1(8-2) (ATCC 37110), pdBPV-MMTneo(342-12) (ATCC 37224), pRSVgpt (ATCC 37199), pRSVneo (ATCC 37198), and pSV2-dhfr (ATCC 37146). Following expression in recombinant cells, CYB5RP can be purified by conventional techniques to a level that is substantially free from other proteins. A description of vectors that can be used to express CYB5RP can be found in, e.g., Goeddel, ed., 1990, Meth. Enzymol. vol. 185 or Perbal, 1988, A Practical Guide to Molecular Cloning, John Wiley and Sons, Inc.

The present invention includes CYB5RP protein substantially free from other proteins. The amino acid sequence of the full-length CYB5RP protein is shown in Figure 3 as SEQ.ID.NO.:3. Thus, the present invention includes CYB5RP protein substantially free from other proteins having the amino acid sequence
5 SEQ.ID.NO.:3. Also included in the present invention is a CYB5RP protein that is produced from an alternatively spliced CYB5RP mRNA where the protein has the amino acid sequence of SEQ.ID.NO.:3 with the exception that amino acids 317-328 are missing.

As with many proteins, it is possible to modify many of the amino
10 acids of CYB5RP and still retain substantially the same biological activity as the original protein. Thus, the present invention includes modified CYB5RP proteins which have amino acid deletions, additions, or substitutions but that still retain substantially the same biological activity as CYB5RP. It is generally accepted that single amino acid substitutions do not usually alter the biological activity of a protein
15 (see, e.g., Molecular Biology of the Gene, Watson *et al.*, 1987, Fourth Ed., The Benjamin/Cummings Publishing Co., Inc., page 226; and Cunningham & Wells, 1989, Science 244:1081-1085). Accordingly, the present invention includes polypeptides where one amino acid substitution has been made in SEQ.ID.NO.:3 wherein the polypeptides still retain substantially the same biological activity as
20 CYB5RP. The present invention also includes polypeptides where two or more amino acid substitutions have been made in SEQ.ID.NO.:3 wherein the polypeptides still retain substantially the same biological activity as CYB5RP. In particular, the present invention includes embodiments where the above-described substitutions are conservative substitutions. In particular, the present invention includes embodiments
25 where the above-described substitutions do not occur in the His boxes of CYB5RP. In particular, the present invention includes embodiments where the above-described substitutions do not occur in positions where the amino acid present in those positions in CYB5RP is the same as the amino acid present in the corresponding position of the sunflower protein depicted in Figure 1 of Sperling et al., 1995, Eur. J. Biochem.
30 232:798-805 when these two proteins are aligned by BLASTP analysis. In particular, the present invention includes embodiments where the above-described substitutions do not occur in positions where the amino acid present in those positions in CYB5RP is the same as the amino acid present in the corresponding position of the

CCCTCTACCCCTGTCCCATCAGGC (SEQ.ID.NO.:15)

One of skill in the art would recognize that many other primer pairs based upon SEQ.ID.NO.:2 would also be suitable.

PCR reactions can be carried out with a variety of thermostable enzymes including but not limited to AmpliTaq, AmpliTaq Gold, or Vent polymerase. For AmpliTaq, reactions can be carried out in 10 mM Tris-Cl, pH 8.3, 2.0 mM MgCl₂, 200 μM for each dNTP, 50 mM KCl, 0.2 μM for each primer, 10 ng of DNA template, 0.05 units/μl of AmpliTaq. The reactions are heated at 95°C for 3 minutes and then cycled 35 times using the cycling parameters of 95°C, 20 seconds, 62°C, 20 seconds, 72°C, 3 minutes. In addition to these conditions, a variety of suitable PCR protocols can be found in PCR Primer, A Laboratory Manual, edited by C.W. Dieffenbach and G.S. Dveksler, 1995, Cold Spring Harbor Laboratory Press; or PCR Protocols: A Guide to Methods and Applications, Michael *et al.*, eds., 1990, Academic Press.

A suitable cDNA library from which a clone encoding CYB5RP can be isolated would be Human Retina 5'-stretch cDNA library in lambda gt10 or lambda gt11 vectors (catalog numbers HL1143a and HL1132b, Clontech, Palo Alto, CA). The primary clones of such a library can be subdivided into pools with each pool containing approximately 20,000 clones and each pool can be amplified separately.

By this method, a cDNA fragment encoding an open reading frame of either 445 amino acids (SEQ.ID.NO.:3) or an open reading frame of 433 amino acids (SEQ.ID.NO.:3 lacking the amino acids at positions 317-328) can be obtained. This cDNA fragment can be cloned into a suitable cloning vector or expression vector. For example, the fragment can be cloned into the mammalian expression vector pcDNA3.1 (Invitrogen, San Diego, CA). CYB5RP protein can then be produced by transferring an expression vector encoding CYB5RP or portions thereof into a suitable host cell and growing the host cell under appropriate conditions. CYB5RP protein can then be isolated by methods well known in the art.

As an alternative to the above-described PCR method, a cDNA clone encoding CYB5RP can be isolated from a cDNA library using as a probe oligonucleotides specific for CYB5RP and methods well known in the art for screening cDNA libraries with oligonucleotide probes. Such methods are described

in, *e.g.*, Sambrook *et al.*, 1989, Molecular Cloning: A Laboratory Manual; Cold Spring Harbor Laboratory, Cold Spring Harbor, New York; Glover, D.M. (ed.), 1985, DNA Cloning: A Practical Approach, MRL Press, Ltd., Oxford, U.K., Vol. I, II. Oligonucleotides that are specific for CYB5RP and that can be used to screen cDNA libraries can be readily designed based upon the cDNA sequence of CYB5RP shown in SEQ.ID.NO.:2 and can be synthesized by methods well-known in the art.

Genomic clones containing the CYB5RP gene can be obtained from commercially available human PAC or BAC libraries available from Research Genetics, Huntsville, AL. PAC clones containing the CYB5RP gene (*e.g.*, PAC clones 759J12, 756B3, 519O13, and 466A11) are commercially available from Research Genetics, Huntsville, AL (Catalog number for individual PAC clones is RPCI.C). Alternatively, one may prepare genomic libraries, especially in P1 artificial chromosome vectors, from which genomic clones containing the CYB5RP can be isolated, using probes based upon the CYB5RP sequences disclosed herein. Methods of preparing such libraries are known in the art (Ioannou *et al.*, 1994, *Nature Genet.* 6:84-89).

The present invention also provides oligonucleotide probes, based upon SEQ.ID.NO.:2 that can be used to determine the level of CYB5RP RNA in a sample. In particular, the present invention includes DNA oligonucleotides comprising at least 18 contiguous nucleotides of SEQ.ID.NO.:2. Also provided by the present invention are corresponding RNA oligonucleotides. The DNA or RNA oligonucleotide probes can be packaged in kits.

In addition to the utilities described above, the present invention makes possible the recombinant expression of the CYB5RP protein in various cell types. In particular, it is advantageous to recombinantly express CYB5RP in plant cells. Such expression in plant cells provides a method for the production of high levels of valuable EFAs such as GLA and OTA in the recombinant plant cells. An example of such recombinant expression of a delta 6 fatty acid desaturase, in that case from borage, is described in Sayanova *et al.*, 1997, *Proc. Natl. Acad. Sci. USA* 94:4211-4216 (Sayanova). The recombinant expression of the borage delta 6 desaturase led to the production of high levels of GLA and OTA in the leaves of the tobacco plants in which it was expressed. The procedures described in Sayanova can be easily adapted to express CYB5RP in tobacco, thus providing an additional, useful way to produce

large amounts of valuable EFAs. Known methods of recombinantly expressing genes in other plant species beside tobacco can be used to express CYB5RP in those other species.

5 The present invention also makes possible the development of assays which measure the biological activity of the CYB5RP protein. Such assays using recombinantly expressed CYB5RP protein are especially of interest.

Assays for CYB5RP protein activity can be used to screen libraries of compounds or other sources of compounds to identify compounds that are activators or inhibitors of the activity of CYB5RP protein. Such identified compounds can
10 serve as "leads" for the development of pharmaceuticals that can be used to modulate the activity of CYB5RP in patients suffering from conditions where that activity is abnormal, *e.g.*, skin diseases, diabetic complications, inflammatory and autoimmune disorders, cardiovascular disorders, complications of viral infection, and retinal dysfunction such as macular degeneration.

15 Such assays may comprise:

- (a) recombinantly expressing CYB5RP protein in a host cell;
- (b) measuring the biological activity of the recombinantly
20 expressed CYB5RP protein in the presence and in the absence of a substance suspected of being an activator or an inhibitor of CYB5RP protein;

where a change in the biological activity of the recombinantly
25 expressed CYB5RP protein in the presence as compared to the absence of the substance indicates that the substance is an activator or an inhibitor of CYB5RP protein.

In particular embodiments, the biological activity of the recombinantly
25 expressed CYB5RP protein is the ability to introduce a double bond into the 6 position of linoleic acid or alpha-linoleic acid.

In some embodiments, it may be advantageous to insert additional steps between steps (a) and (b). Such additional steps might include lysing the host cell and fractionating its contents in order to partially purify the recombinantly
30 expressed CYB5RP, thus facilitating exposure of the recombinantly expressed CYB5RP to the substance as well as to any substrate used in the assay.

The present invention includes activators and inhibitors identified by the methods described herein as well as pharmaceutical compositions comprising

such activators and inhibitors. The activators and inhibitors are generally combined with pharmaceutically acceptable carriers before use to form pharmaceutical compositions. Examples of such carriers and methods of formulation of pharmaceutical compositions containing activators or inhibitors and carriers can be
5 found in Remington's Pharmaceutical Sciences. To form a pharmaceutically acceptable composition suitable for effective administration, such compositions will contain an effective amount of the activator or inhibitor.

Therapeutic or prophylactic compositions are administered to an individual in amounts sufficient to treat or prevent conditions where CYB5RP activity
10 is abnormal. The effective amount can vary according to a variety of factors such as the individual's condition, weight, sex and age. Other factors include the mode of administration. The appropriate amount can be determined by a skilled physician.

Compositions can be used alone at appropriate dosages. Alternatively, co-administration or sequential administration of other agents can be desirable.

15 The compositions can be administered in a wide variety of therapeutic dosage forms in conventional vehicles for administration. For example, the compositions can be administered in such oral dosage forms as tablets, capsules (each including timed release and sustained release formulations), pills, powders, granules, elixirs, tinctures, solutions, suspensions, syrups and emulsions, or by injection.

20 Likewise, they can also be administered in intravenous (both bolus and infusion), intraperitoneal, subcutaneous, topical with or without occlusion, or intramuscular form, all using forms well known to those of ordinary skill in the pharmaceutical arts.

Advantageously, compositions can be administered in a single daily dose, or the total daily dosage can be administered in divided doses of two, three or
25 four times daily. Furthermore, compositions can be administered in intranasal form via topical use of suitable intranasal vehicles, or via transdermal routes, using those forms of transdermal skin patches well known to those of ordinary skill in that art. To be administered in the form of a transdermal delivery system, the dosage administration will, of course, be continuous rather than intermittent throughout the
30 dosage regimen.

The dosage regimen utilizing the compositions is selected in accordance with a variety of factors including type, species, age, weight, sex and medical condition of the patient; the severity of the condition to be treated; the route

of administration; the renal, hepatic and cardiovascular function of the patient; and the particular composition thereof employed. A physician or veterinarian of ordinary skill can readily determine and prescribe the effective amount of the composition required to prevent, counter or arrest the progress of the condition. Optimal precision
5 in achieving concentrations of composition within the range that yields efficacy without toxicity requires a regimen based on the kinetics of the composition's availability to target sites. This involves a consideration of the distribution, equilibrium, and elimination of a composition.

The present invention also includes antibodies to the CYB5RP protein.
10 Such antibodies may be polyclonal antibodies or monoclonal antibodies. The antibodies of the present invention are raised against the entire CYB5RP protein or against suitable antigenic fragments of the protein that are coupled to suitable carriers, *e.g.*, serum albumin or keyhole limpet hemocyanin, by methods well known in the art. Methods of identifying suitable antigenic fragments of a protein are known in the art.
15 See, *e.g.*, Hopp & Woods, 1981, Proc. Natl. Acad. Sci. USA 78:3824-3828; and Jameson & Wolf, 1988, CABIOS (Computer Applications in the Biosciences) 4:181-186.

For the production of polyclonal antibodies, CYB5RP protein or an antigenic fragment, coupled to a suitable carrier, is injected on a periodic basis into an
20 appropriate non-human host animal such as, *e.g.*, rabbits, sheep, goats, rats, mice. The animals are bled periodically and sera obtained are tested for the presence of antibodies to the injected antigen. The injections can be intramuscular, intraperitoneal, subcutaneous, and the like, and can be accompanied with adjuvant.

For the production of monoclonal antibodies, CYB5RP protein or an
25 antigenic fragment, coupled to a suitable carrier, is injected into an appropriate non-human host animal as above for the production of polyclonal antibodies. In the case of monoclonal antibodies, the animal is generally a mouse. The animal's spleen cells are then immortalized, often by fusion with a myeloma cell, as described in Kohler & Milstein, 1975, Nature 256:495-497. For a fuller description of the production of
30 monoclonal antibodies, see Antibodies: A Laboratory Manual, Harlow & Lane, eds., Cold Spring Harbor Laboratory Press, 1988.

Gene therapy may be used to introduce CYB5RP polypeptides into the cells of target organs, *e.g.*, the pigmented epithelium of the retina or other parts of the

retina. Nucleotides encoding CYB5RP polypeptides can be ligated into viral vectors which mediate transfer of the nucleotides by infection of recipient cells. Suitable viral vectors include retrovirus, adenovirus, adeno-associated virus, herpes virus, vaccinia virus, and polio virus based vectors. Alternatively, nucleotides encoding

5 CYB5RP polypeptides can be transferred into cells for gene therapy by non-viral techniques including receptor-mediated targeted transfer using ligand-nucleotide conjugates, lipofection, membrane fusion, or direct microinjection. These procedures and variations thereof are suitable for *ex vivo* as well as *in vivo* gene therapy. Gene therapy with CYB5RP polypeptides will be particularly useful for the treatment of

10 diseases where it is beneficial to elevate CYB5RP activity.

The present invention is not to be limited in scope by the specific embodiments described herein. Indeed, various modifications of the invention in addition to those described herein will become apparent to those skilled in the art from the foregoing description. Such modifications are intended to fall within the

15 scope of the appended claims.

Various publications are cited herein, the disclosures of which are incorporated by reference in their entireties.

WHAT IS CLAIMED:

1. A recombinant DNA molecule encoding a polypeptide having the amino acid sequence of SEQ.ID.NO.:3.
5
2. A recombinant DNA molecule comprising a nucleotide sequence selected from the group consisting of:
SEQ.ID.NO.:1;
SEQ.ID.NO.:2;
10 SEQ.ID.NO.:2 lacking positions 1,019-1,054;
positions 71-1,405 of SEQ.ID.NO.:2; and
positions 71-1,405 of SEQ.ID.NO.:2 lacking positions 1,019-1,054.
3. A DNA molecule that hybridizes under stringent conditions to
15 the DNA molecule of claim 2.
4. An expression vector comprising the DNA of
claim 1.
- 20 5. A recombinant host cell comprising the DNA of claim 1.
6. A CYB5RP protein, substantially free from other proteins, having an amino acid sequence selected from the group consisting of SEQ.ID.NO.:3 and SEQ.ID.NO.:3 lacking positions 317-328.
25
7. The CYB5RP protein of claim 6 containing a single amino acid substitution.
8. The CYB5RP protein of claim 7 where the substitution is a
30 conservative substitution.
9. The CYB5RP protein of claim 6 containing amino acid substitutions where the substitutions do not occur in positions where the amino acid

present in CYB5RP at those positions is also present in the corresponding position in the delta 6 desaturase from sunflower when CYB5RP and the delta 6 desaturase from sunflower are aligned by BLASTP analysis or where the substitutions do not occur in positions where the amino acid present in CYB5RP at those positions is also present in the corresponding position in the delta 6 desaturase from *Synechocystis* when CYB5RP and the delta 6 desaturase from *Synechocystis* are aligned by BLASTP analysis or where the substitutions do not occur in positions where the amino acid present in CYB5RP at those positions is also present in the corresponding position in the delta 6 desaturase from borage when CYB5RP and the delta 6 desaturase from borage are aligned by BLASTP analysis.

10. An antibody that binds specifically to the CYB5RP protein of claim 6.
11. A DNA or RNA oligonucleotide probe comprising at least 18 contiguous nucleotides of at least one of the sequences of claim 2.
12. A method for determining whether a substance is an activator or an inhibitor of CYB5RP protein comprising:
 - (a) recombinantly expressing the CYB5RP protein of claim 6 in a host cell;
 - (b) measuring the biological activity of the recombinantly expressed CYB5RP protein in the presence and in the absence of a substance suspected of being an activator or an inhibitor of CYB5RP protein;where a change in the biological activity of the recombinantly expressed CYB5RP protein in the presence as compared to the absence of the substance indicates that the substance is an activator or an inhibitor of CYB5RP protein.
13. The method of claim 12 where the biological activity of CYB5RP protein is the ability to introduce a double bond into the 6 position of linoleic acid.

14. A pharmaceutical composition comprising an activator or an inhibitor of CYB5RP.

5 15. A method of treating macular degeneration comprising administering to a patient an effective amount of the pharmaceutical composition of claim 14.

1/19

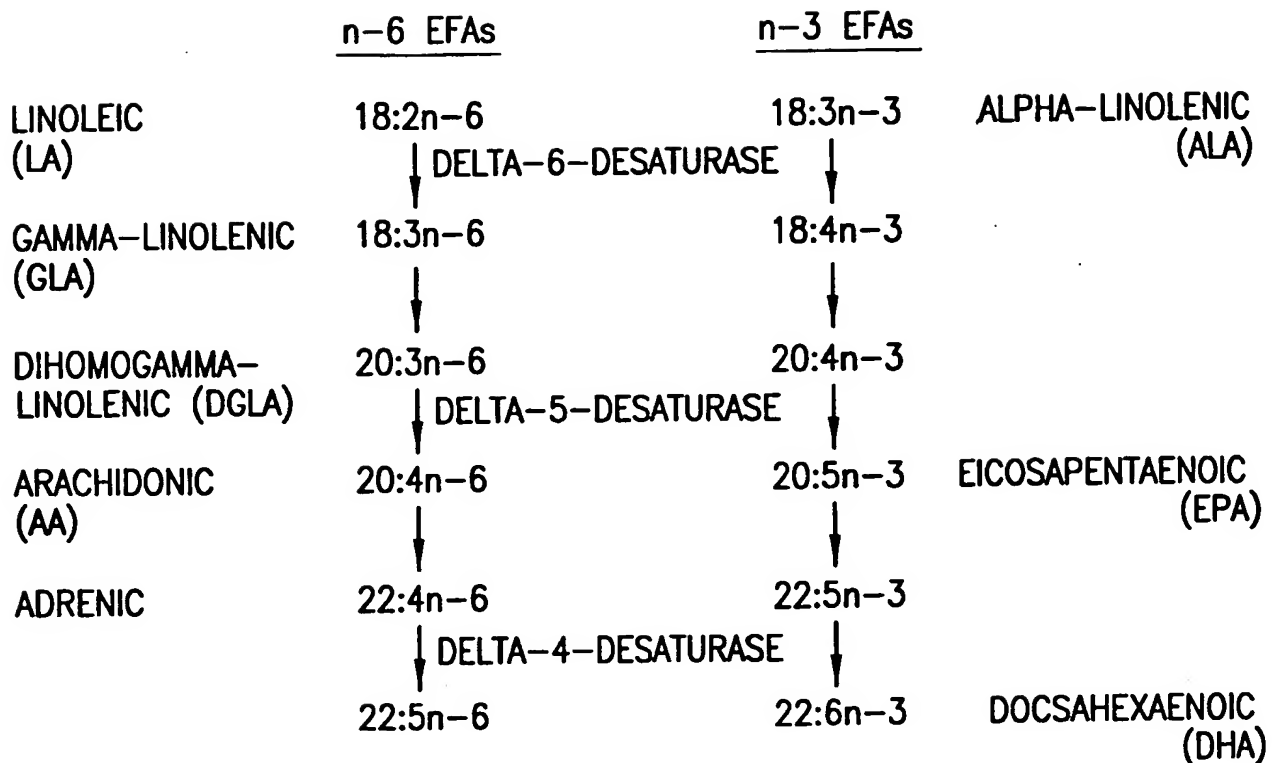


FIG.1

2/19

```

1   gctcacagac cgggactccg cctccggttc ccgagggcgt ggcgagggcg
51  tgcgggacgc ccaacaggtg cgtgttgtgt ccccaggccc cgcgctccgg
101 gtggagtcaa gagcctggaa gccggcagcc cgggaaaagg gggcgggacg
151 gtgccccggg gcagggctgg gtggcgggcg ctgtcctccc gggagggggc
201 ggccgcctcg acgccgccct ccctggcggc caatggagac cgaggccccg
251 cgcctggatt ggagcggacg cgggggtcag ccagccttgg gggccggggc
301 ctggccgggg gcgggggggc aggcgaggcg aggcgggcgc cgtccgcgcg
351 gttataaagg ggggagttcc ctgcgccgcg agccgggagg cgcacgctcg
401 ctcgtaacgc gcccgcgggc gcaggggcgg gccggagcag cgggcggcgg
451 cggaggcggc gcccggggag gctCTTCGCT TCCCTCGGGG TCTTGCTCGG
501 ACCTCGGCCA CCGCCTGGGA TCCCCAGGAC TCGTGCGTGC AGCATGGGCG
551 GCGTCGGGGA GCCGGGACCG CGGGAGGGAC CCGCGCAGCC GGGGGCGCCG
601 CTGCCACCTT TCTGCTGGGA GCAGATCCGC GCGCAGACC AGCCCGGCGA
651 CAAGTGGCTG GTCATCGAGC GCCGCGTCTA CGACATCAGC CGCTGGGCAC
701 AGCGGCACCC AGGGGGCAGC CGCCTCATCG GCCACCACGG CGCTGAGGAC
751 GCCACGgtaa ggaagccata aggaagccac ccaccggcgg gtggagcctg
801 gagctcggtc gtgggcgtga tgtcccgcct cacctgtggg gccttagcat
851 cctccctccc ctcgctgacc tttgacctcc acgccgggac ccagagttgg
901 ggtggactag ccagggccag atgtggggta gggagggcag ttccttgcgt
951 ggaggacccg cagctgtcca cggagcaggt ctgcggggga ggagggggcc
1001 tcagaggtgg gtgtgtcatg ctgcagagcc tgccctgggt gaggggctgc
1051 cctgttgctc ccaggtccct gtttcagttc tgggtcccca tgctgggtgc
1101 ttgctgagtg ctaggggtag ggcagggcag ggtccccagg ggccggtaag
1151 gacatgccat tagaggctgg gggctggggc ggctgaggt ctgtggcttt
1201 cccaagagct tctgtaaagg gctcagggac agtgactcac ctctccgggc
1251 tagagctgc acgtgggagg gctttgccag ccaggctggg tgggcctctc
1301 ctggaagcac agtcacccca ggaacaggct ggcccttggg gaccccaact
1351 tcccaatccc agccctgtc tagacaggca gggatgtagc ctggccccag
1401 ggtactgtct ggctggagtc cagtgggtga gcagcccagc cagccccctt
1451 tccttagtta ccacctgca taataggggt tggggccacg atgccctgtc
1501 cttgaccctc caaatttcta ggttggccac actgggtatc aggaaggtct
1551 tcaagaccgg aggacatgaa tctgaatgc tggctttttg ggcagcagcg
1601 gaggttctgt ccagtcaccg gactgtcggc gtccctcttg ccagggccac
1651 ctgctctctg ccgattgcca tctccagcat gttggacaat cttcactgga
1701 ctctttgagg aagaaagccc ctcttttccc tttccacccc atgaagctga
1751 ggagtggaaa taagaatcct cctgaaattc taaaaaaaga aaaaaaaaaa
1801 aaagagaacg ccttgtccgt ggctgttcag gcgccagacg ctggcccag
1851 gggacagcac agccgtggga tgaagcagcc tgggggcagt atttgagcgt
1901 gcaggtgttt gcatgtctgg gtgagtgagg tgtgtgtgcc tgcctttctg
1951 ccagggcgtg gcgaggtgag gggcacggct tctccccaaa ggccttgctg
2001 agccctggcc tcccttcaag gactcttgtg gatgcctgct ctggtctttt
2051 tttaaaaaag tatctatttt atttattatt atttgtttaa aaatagagac
2101 aggtgtctac tatgttgctc gggctggctc caaagtcctg ggttcaagca
2151 ttccctcctg ctcagcctcc gaaagttctg ggattacagg catgagccac
2201 cactcccggc ctgctctagt cttttgtaac cttagaggaca gtatggatac
2251 agaaaacttt actcccacc aaccgccgga gacagagtct tgctctgcca
2301 cccagactgg agtgcaatgg cgccatcttg gctcactgca acctccgctt
2351 cccaggttca agcgattctc ctgcctcagc ctcccagata gctgggatta
2401 cgggcacgcg ccaccacgcc cagcatattg tatttttagt agagacgggg
2451 tttcaccatg ttggccaagc tgggtctcgaa ctccctgacct cgtgatccac
2501 ccacctcggc ctcccaaagt gctgggatta caggcgtgag ccaccacgcc
2551 cggctgggat acagaaagct tttatttcat cactgtttcc tgcctggtgc

```

FIG.2A

3/19

2601	caggcccatg	ctgggggttcc	tcccaagtgg	aattactgac	ttaacattta
2651	gcttgggatc	ctgagacttc	catcacacag	ttttctcatt	gattcgcagc
2701	caataatatc	tgttttataaa	acatctcagg	ccgagcgctg	tggctcacac
2751	ctgtaatccc	agcacttttg	gaggctgagg	tgggcagatc	acctgaggtc
2801	gggagtttga	gaccagcctg	accaacatgg	agaaaccctg	tctcttctaa
2851	aaaaatacaa	aattagccag	gcgtggtggc	gcatgcctgt	aatcccagca
2901	ctttgggagg	ctgaggcagg	agaatcgctt	gaaccagga	gacggaggtt
2951	ccggtgagcc	gagatcgcg	cattgcactc	cagcctgggc	aacaagagca
3001	aaactccgtc	tcaaacaaac	aaacaaaaaa	catctctctg	ctccttgggg
3051	ccgggtgcca	gctctgctat	tggaggcact	gagcgacctt	gaagcaggca
3101	tgtcactcct	ctgtgcccc	gtttactcat	ctgtaaagt	ggagagctgg
3151	ggcagacagt	gagctggctg	agggcaggac	tgtgtctcct	caagcccatg
3201	gccaggggct	gccaggtagt	agtttgtatt	cggtaaatgc	tgctggcccc
3251	taagtgtgag	cgtgccctgc	aaactgcagc	gtatggtggg	acagccctgc
3301	acggctaccc	ctttcctggg	tgaccttatt	tggttacggt	cctatctgaa
3351	gtaggaaagg	gacactttag	gctgtctctt	agctccctca	agggccacaa
3401	gcctggacta	gagttgccag	gaaatcttgg	tccattcagg	ccaaagggac
3451	tgtgaggttg	ctgggatggg	gcaatcagtc	tttgtccatg	atgaaccac
3501	agggtagacc	aggggttggg	ccagcccagt	gccctgtgta	gttgagccca
3551	ggccccaggc	atcccatccc	gggcggtggc	ctcaggtgga	ggtggggcag
3601	ccagttgcca	gggatgtgtt	ccagcggtca	cctctcacca	gccccggctg
3651	cccatcagct	gttctcaagt	ccaggcaatg	aagccttctt	gccaggaaat
3701	tcccagagtt	tctgtgccat	gaagtcagcc	tgtggccatc	ttgggacaca
3751	aggccgggtg	ccctggggag	agtactctgg	gcccttggcc	aggtttgtct
3801	gagagtcata	ggcagcctga	tactagtgga	gccagccagg	gagggatgag
3851	gcccagccgc	tgctggccat	aagtatataa	gggccatgtg	ctgagtgcct
3901	actatgtgcc	aggttttgaa	atcagtactt	gatttattga	aaccctctct
3951	tttaactctc	aaggtgcccc	tatgaggcac	gtaccattta	ttgttattgc
4001	cacttgacag	atgagaaaac	agaggctcag	agaggcaaag	tggcttgaaa
4051	ttcagtgatt	ggtctgggat	ttgaatccac	agccatgttc	ttaagggcat
4101	gctatgctgc	cacctatcct	gtttattttc	ggcactcatt	gattcttcaa
4151	tgtttgactc	attaaatcca	tcagttagca	tcttctctgt	gtcatgcatg
4201	gttctcacct	ctgaagatgt	agctgtgagc	aaaacttcta	cagggaatga
4251	gttcacagca	gagggatcag	ctagagcaaa	ggctcagagg	tgggaccgtg
4301	cgtcctgtgt	tccaggaata	cagtatggct	gcagcagaga	gcagtggaga
4351	gagggcctgg	cagtgaggtc	tagaggcggc	cgggctggct	catgtggat
4401	gtttgtgtcc	tcggaaggac	tttggcttta	ttttaagag	gatggggagc
4451	ccagagagc	acagcaggga	agcctgggga	gtctgatgga	catttaaaag
4501	gatccttaat	ggagagagt	aaggcagagc	cttcagaag	ggtaagagaa
4551	gggaggatgg	agacctgcc	tcccccaagg	gaggccactc	agaagaggta
4601	gagtgtggcc	agggcagaga	gcaagagagg	ctgtggacac	aggcacactg
4651	gtccagttag	agccattaga	cacattagat	ttagcttcat	gttgtcttta
4701	gagagggagc	cagcctggcc	tcgctctatg	atcttggaca	catcctttca
4751	cttctgggtc	tcagtttccc	cattagtgtg	atgaggatga	gaatgctttt
4801	gtcctgggca	cactatgagg	gtggtgctgg	gcacctgggt	gcctggttac
4851	catgggcaac	aaagctctat	tcatgggtgt	ggtgaatgca	ttgcccacag
4901	caactcaggg	cggatgagga	gtttcccagc	agcccttggg	gccctttcgg
4951	ctgaagccct	aacaactgtg	ggaaaatcca	agttccagca	gacccctga
5001	gccctctgcc	ttaggacctt	ccttctaggt	ggttctctga	gcctggcctg
5051	agctggagga	gggagtggcc	agtgtgtcag	cagaggctgc	ttcatagtaa
5101	ttgcagccaa	cagttattga	ctaggcactg	ttctgagggg	tttagatgtg
5151	gtaactgatt	gaattcgcct	aacaacttta	tgaggttaagt	cctattgtta
5201	gcccattttg	tagatgagga	gactgagttt	gaaactgggg	ggtgtaatgg
5251	aaccttctca	ggacccttga	agggtagggc	ctttgtactc	gggccacgag

FIG.2B

4/19

5301	ggtgggggttt	gtgtctgggt	gggagctggg	gagggacagg	actaggatta
5351	ggcagatctg	aggccacagg	agttgggttg	ggggtggctc	cagagccact
5401	ccactccctc	ctaccacatt	gactgccttg	aaagtcccct	aatggccact
5451	cccatgaagt	gtgactgctc	tgggctcccc	gcaggcgttt	tctgcaaggc
5501	caccgcccac	ccaggcccct	tccccagagg	ggctgcagtg	ccttgctcct
5551	tccttgtggg	aagagttggg	attgtctggc	gtcagcagga	tactgcccct
5601	gggcatccct	cccggctctc	tcctgcgggt	ttctgatgaa	acagccaggc
5651	tccagtagtg	gagccagagg	tcagtgggtg	agagaggacc	aggagccaga
5701	gggtatagct	gctttggggc	tactgtgggg	tcagggacac	ttgtgaggcc
5751	aagcgtcctg	gctgcaggag	ccctcacata	tatgccacc	cttcaccagg
5801	acattgaggg	gtgctggggg	acaggggtag	ctttttgggg	gtgtctgcct
5851	tcgacttggg	ctccgctaca	caggccaaat	ttggatgtcc	catgtttaga
5901	gctgtgtttc	tttgggacct	cttggggcct	cagtttcctc	atctgtaaaa
5951	tgggatactg	atagtgtctc	cccactggcc	tcctctgacg	ggcgccaggg
6001	agaggatggg	acggagcatg	gtgtgctggg	cacgctcctg	ctgtaccac
6051	ccacctggga	gaggggagag	gcaggaatgt	cctgggggtg	tcctttgagg
6101	catagccctg	tcaccccaac	atcctacaaa	ggcatgagaa	ggcagcgagg
6151	acagaccccc	accacctgag	ccctcagcag	ccctgccaca	ctccctgctt
6201	caccccttcc	ctgactgata	tggcacattc	ttgattctcc	tagggagtga
6251	cccaaaatcc	ctccctgccc	tgtgtgtctc	ctgggggtga	aggaggctgc
6301	cagccctccc	tctctcccag	cctcaggctt	ggccaggact	taacaggcag
6351	gcagagaagc	agcttctcca	ctctcttccc	tgacacctgt	aggccctccc
6401	tgcaggcact	tacctctaag	tggactctca	ggaggaggct	catcagggct
6451	gcagggctca	gaaagagctg	ggctgtggag	ctcttgccaa	ccgccaggcc
6501	ccttctaagt	gctttagcgc	caccgactgc	atcctcccag	cagccttgtg
6551	agatggggat	ttgtggttcc	cagtttactg	atgagaaata	ctgatgagag
6601	atgggtgtgg	tcttgtctgg	ggctccctgg	ctcctggata	gcagctcagg
6651	ttccatcctg	ggcaggctgg	ctctgggaca	ccccccgac	cagctgctgt
6701	gtgggattca	cgggtggggc	tggggcaggc	gtgggatcct	ggggccaact
6751	gagccactct	aggcttccag	ggaccaaggc	caggctgagc	tgtctctgta
6801	tcctgagaga	gcataaacat	cacagaagat	gggcccgggt	tcgaatccca
6851	gctctgccac	tactaactgg	gacctgggca	ggggctccct	cccgtgagc
6901	cttcatttcc	tcaccagcaa	aatggttcgt	gcccctgctt	tgggggctgt
6951	ggagggttgg	ctcttgtcta	cttgttcata	cctgctgttg	agcagctgct
7001	ctgtgccggc	ctctgaggat	gccactgtga	acagagcctg	tcgctacctc
7051	caggagcttg	tgtttagggg	tggcgttttg	attccagcac	tttccaccag
7101	ctctgctccg	gtacccgatg	agagacgtcg	agtgccgctt	tccactcgct
7151	tgggtgcgtg	tgggggttgg	ggggacaggg	ctttgtgcac	gtagccctgg
7201	gtggatgttc	ctgggtgcac	ttaggggtgtg	tgagggtggg	acctcccaca
7251	gttccctgag	gctccactga	tgagggtccaa	gaaccgcctt	cctgcccccc
7301	agcccaggct	cccagcagct	gggcccttgg	cttcttgaga	tagtgactgg
7351	cctcacggca	aggacccccg	cacaccacct	aggagaactg	ctgcttcccc
7401	tctgttccag	gagtggcgac	aagcacagtt	tttcgctttt	gtttttgttt
7451	tcttcacttt	aagttccggg	aaacgtgcag	aatgtgcagg	tttgttacat
7501	aggtatacat	gtgccatggg	ggtttgctgc	accgctcaac	ccctcatcta
7551	ggttttaagc	tccatataca	ttaggcattt	gtcctaatac	tctccctccc
7601	cttgcccttc	acccgcccag	taagccccgg	tgtgtgatgt	tcccttccct
7651	gtgtccatgt	gttctcattg	ttcaactctc	acttatgagt	gagaagagac
7701	ctggactctg	atctaacctc	gggtcaaatg	aactgtgtga	ccttgaagaa
7751	gtagcttaac	ctctctgagt	cttagcttct	gcctggcacc	cccatcctta
7801	aggagaggcc	cacagaggac	caggtcacat	gacctcagcc	agttccagag
7851	aaggctgttt	gcttccaggt	ttcggcctga	gtccaggccc	ctgccctact
7901	cgcactccct	gatagcatga	gaagcacagc	cccagggtgc	ccaccagct
7951	ctgagagccc	agcctgcttc	ccagggaact	gtcacagccc	cacctgtccc

FIG.2C

5/19

8001	ttccccagct	ggagccctgt	caatggcttt	ggggttctct	gacacagccc
8051	tgagggggct	cacacttccc	cttatcattg	caaggggtag	atctggcttg
8101	aaggccctgg	ggcaggcttg	gttctgtcct	cccctgtcag	tgcctcgaca
8151	gggctggcct	gggtgaatca	ggaccaacgg	gaaaggaggc	gaggagacca
8201	atctggaccc	aagatcctca	gctcaataag	gtggccccag	aactgacatg
8251	gggtgataga	gggaagggct	gggagggagg	agattcttggg	gccgcagcca
8301	cagcttgcac	gttgcgccgg	gtgtgtctgt	gcgtgccagc	tgcattcttg
8351	cgtaccatgt	gtgcaaggct	gtgtttggct	gagtgttcat	gtgggccgtg
8401	attgtgggca	tgtttctgag	tgtctgagtg	atgcctgctg	gtgtgggctg
8451	gtgggtgtgt	ctgcatgtgc	gtgtgtgtct	ggggagtttc	aaaggagaaa
8501	gagggactca	ccatcacgct	ggctcagcct	taaaaaggta	ggacatcctg
8551	acacgtgctg	caacatggat	ggaccttaag	gacattgtgc	tgagtgaaac
8601	aagccagagg	caaaggaaca	aacatgtgat	ttctcccaga	tgaggtttcc
8651	ggaggaggca	gatctgtatg	gacagaaggt	agcatgggtg	ttgccggggc
8701	agggggagga	gagaatggag	aattagtggt	taatggggac	agagtttcag
8751	ttgggggaagg	tgaaaagggt	ctggagctgg	atgatgggtg	tggttggaca
8801	acactgtgca	tgcacttaat	accactgagc	tggacaccta	aaaatgctta
8851	caatggtaaa	tttcatgtat	attttactac	aatttttaaa	aaattggctg
8901	ggcgtgggtg	cttatgcctg	taatcccaac	actttgggag	gccaaggcgg
8951	gaggattgct	tgagctcagg	agttcaaacac	cagcctgggc	aatatgggtg
9001	aaccccgact	ctacgaaata	tacaaaaatt	agcctgggtg	ggtggcctgc
9051	acctctaata	ccacctaact	agtaggctaa	ggcacaagaa	tctcttgaac
9101	ctggggagggtg	gaggttgcag	taagccgaga	tcatgccact	gcaacccagt
9151	ctgggcgaca	gagcaagact	ctgtctcaaa	aaataaaaaga	taaataaaaa
9201	aattagaggc	caggtgtggc	tcacacctgt	actctcaaca	ctttgggagg
9251	ctgaggtggg	aggatcgctt	gaagtcaaggc	atttaagaca	tgccataggca
9301	acatagttag	accttgactc	tacaaaaaaa	ttcaaaagtt	aatgagacat
9351	ggtggcatgt	gcctgtagtc	ctagctgctg	gggaggctga	ggtgggagga
9401	tcacttacga	ccaggatttc	aaggctgcag	tgagctgtga	ttgcatcact
9451	gcactccagc	ctgggtgacag	agtgaaggcc	tgtctcaaaa	aaatttttca
9501	gtgtttttct	gggctggggc	tggtggctca	ttcctgtaat	tccagcactt
9551	tgggaggctg	aggtgggtgg	attgcttgag	cccaggagtt	taagaccagc
9601	tgggcaacat	ggcaaacctc	atctctacaa	aaaataaaaa	taaaaaatta
9651	gctgggcatg	gtggtgcaca	cctgtactaa	cagctacgag	agaggctaag
9701	gtgggaggat	cacctgagcc	cgggaggttg	aggctgcagt	gagccatgat
9751	tgcaccactg	cactctagcc	tgggcgatac	agcaagaccc	tatctcaaaa
9801	aaaaaaaaaa	aaaaaaaaaa	aaaaacaccc	agtggggtca	gtagaacccc
9851	aagagtcttc	ttccctccca	gctcccctgt	acaccagccc	cagctctgca
9901	ggtagctggg	ggcccagaca	gcttctctggg	gaccccagc	cttccctctg
9951	cccttttttc	taccagtttt	gctgcccctc	cttcaagact	catgtccaga
10001	gggggtgaga	tctgcactta	tacagccccc	tcctctgtaa	tgagttagcc
10051	aagtcagccc	aggttatttc	agaaggggca	ccctaccagc	ccccagtc
10101	ccaagctgcc	ctgggcctat	aaaagcaggc	aaggggaccc	ctagtagatc
10151	atgtagggtg	tacctcttag	tgggtgctgg	aggggcctga	agtgtttct
10201	tccccaggg	tggtaggaga	atgtcctggc	agtgacttca	gggcccgtg
10251	tcacttccgt	tttaagactc	accagctggg	aggctcatta	gcaagaggac
10301	aataggaggc	ccctgtcctc	agtcagcttt	cttcaaaggt	gtttccttta
10351	gcaactggga	ggcctccctt	ctccagaccc	atggggacaa	caccacccag
10401	ctactgggtc	tataagctgc	tgtatggctc	tggctagccc	attcagagaa
10451	agcctctgaa	agtacaagga	aaaaaatcag	tccaagagct	gtgaacaatt
10501	agtgaagcga	ttacaatacc	aagaccacag	gcagacctgg	aaggctaagt
10551	gagcccaggt	gtgaagttca	agcttacttt	acttctgggc	cacttctctg
10601	ctggtctctt	tccttgcccc	ttatctttct	cctgggtctgt	cttctcttct
10651	cacccccttt	ctttactctt	tcttcttctt	cctgcatcgt	actccacccc

FIG.2D

6/19

10701	cactccagct	attacacaga	atcgcgagaa	tggttgatta	ttcattttat
10751	ttatgatgtt	ttcttttttg	taaaaataga	gacaaggctc	cactatgtgg
10801	cccaggctgg	tcttgaactc	ctggcctcaa	gcaatcctcg	tgctttggcc
10851	tcttacagtg	ctgggattac	agatgtgagc	caccatgcct	ggcccathtt
10901	atttacttta	aaaaaaaaat	taggctgggc	gcggtggctc	acacctataa
10951	ttccagcact	ttgggaggcc	aagggtggga	gatcaactga	gggtcaggagt
11001	taaagaccag	cttggccacc	tggggtcagg	agtttgagac	cagctactcc
11051	ggaggctgag	accggagaat	tgcttgaacc	caggaggtag	agggtgcaat
11101	gaactgagat	catgccattg	catgccagcc	tgggcaacag	agcaagactg
11151	tctcaaaaaa	aaaaaaaaatt	atgttttgtg	ctcctgcttc	ctgctttgta
11201	agtcaaataca	gtttaactgt	tcaagtgtct	tccttgcaaaa	cccccaagga
11251	ctcaatgtgt	gtcgcccttg	actgatcccc	ccgccccgtg	acccagtggt
11301	cctcagttcc	aggtttttccc	acctaccctt	caccactgc	ttatgtttat
11351	aaaaacgggg	taaatcaaat	gttcgtgacc	cagatcttat	tctacatgca
11401	gtggaaactt	gtatgactta	agcttttttg	aaaagcagaa	ccttttttcg
11451	tggttcaaga	aatcaaagtc	ttccccggag	gtctttctgt	aaatccagag
11501	ctgcagatgt	ttgaccgtgt	tcagagaggg	gcccttgtgc	tgggtgaagt
11551	ggatggggca	cagcaggcaa	tgggtgaaaa	gcaggacaac	ctggggccct
11601	gggaggacca	gggaggggccc	atgtctttga	ctgttcatca	gccggctgac
11651	ttcctgtccg	cctgtcgtct	gctctgccc	tccatccgta	gtccttccgc
11701	ctgtctctgc	tggttgccgc	tgtgtactc	agctgtgtct	gtcgttccgc
11751	ctgactgtct	gctctccttc	agGATGCCTT	CCGTGCCTTC	CATCAAGATC
11801	TCAATTTTGT	GCGCAAGTTC	CTACAGCCCC	TGTTGATTGG	AGAGCTGGCT
11851	CCGGAAGAAC	CCAGCCAGGA	TGGACCCCTG	AATgtgagcc	agagccctag
11901	gagaggctca	gcccctgagg	gagggggatg	gctggaggggc	tgggagacat
11951	tgccacatgg	ccaggagcag	ctccctcggc	attcgcccaa	ggggatgcag
12001	agccagggtc	gagcctgccc	tcccctccca	gggggcaggc	agttgaaagt
12051	gaagctgtag	ggatgccctg	agaagtccag	ggctccagat	ctggtttagc
12101	caggcactcg	tttggatccc	gaggcaagct	ccctccctgt	tgtcgcccag
12151	tgtcccccac	aaaaggagga	ttttgatgaa	ctgatttctc	tcctggctgt
12201	agcgtcttac	ccaccccata	cctttttggga	gggagaggag	gcttcaccac
12251	cagccagtgc	tccagctcac	accccgggct	gggtactctt	gtcacttcat
12301	tcctctttgc	ccacaccctt	tgggcctggc	gatgggagga	gcggctgggg
12351	ctccaggaga	atgggggttg	ggaggaattt	cttccttggc	tgatcgggcc
12401	ctctgctatg	gcagGCGCAG	CTGGTCGAGG	ACTTCCGAGC	CCTGCACCAG
12451	GCAGCCGAGG	ACATGAAGCT	GTTTGATGCC	AGTCCCACCT	TCTTTGCTTT
12501	CCTACTGGGC	CACATCCTGG	CCATGGAGGT	GCTGGCCTGG	CTCCTTATCT
12551	ACCTCCTGGG	TCCTGGCTGG	GTGCCCAGTG	CCCTGGCCGC	CTTCATCCTG
12601	GCCATCTCTC	AGgtgacccc	agttctgtgt	tgcagccacc	ttaactgccc
12651	aacagacgtg	ggcccccatg	catctgggca	ttgtgaacat	atttgctaaa
12701	tgaatgaatg	gacctatgaa	aggatgaatg	gatgaataaa	cagatgaatg
12751	agtgaacagt	ctgaaggccc	atcaggcatg	tctgtgggtc	aagctgcatt
12801	ccagatgagc	caagaagttc	cttcttgaac	agattccgat	caagcacagg
12851	gccactgagc	cagaggctgc	tgccctgcag	cttcatgaca	cttacgagcc
12901	cctccacctc	cctgggactc	agttctcatc	tgtaaaaaga	ggacactggc
12951	ccacaagggt	cttgaaatgg	agcattagca	cgggggtacc	ctgcaagctg
13001	aaaggattca	ctggggcccc	aggccctggc	gggctccgtc	cttcccaaca
13051	gcttctgacc	ctgcctctct	ccccagGCTC	AGTCCTGGTG	TCTGCAGCAT
13101	GACCTGGGCC	ATGCCTCCAT	CTTCAAGAAG	TCCTGGTGGA	ACCACGTGGC
13151	CCAGAAGTTC	GTGATGGGGC	AGCTAAAGgt	gaggggtgggg	tgggtggtca
13201	gccaggtgct	gggtggcgtc	gggtctgccc	aagtgtgtgg	gcacagtcgg
13251	gggcacagcc	tgccctgaga	gccccctcct	cctccacagG	GCTTCTCCGC

FIG.2E

7/19

13301	<u>CCACTGGTGG</u>	<u>AACTTCCGCC</u>	<u>ACTTCCAGCA</u>	<u>CCACGCCAAG</u>	<u>CCCAACATCT</u>
13351	<u>TCCACAAAGA</u>	<u>CCCAGACGTG</u>	<u>ACGGTGGCGC</u>	<u>CCGTCTTCCT</u>	<u>CCTGGGGGAG</u>
13401	<u>TCATCCGTCG</u>	<u>AGgtgggtgg</u>	<u>ggagggacct</u>	<u>ggacaacctc</u>	<u>tggctggggc</u>
13451	<u>tgcagctgag</u>	<u>ggggagctaa</u>	<u>tgcactgggt</u>	<u>ccccactctg</u>	<u>ccctgacct</u>
13501	<u>agccccctgat</u>	<u>ctggcctcca</u>	<u>ctctggctgg</u>	<u>gccaagctct</u>	<u>gccccctgt</u>
13551	<u>ctttccttcc</u>	<u>cacctcccaa</u>	<u>cctgctgggg</u>	<u>acgaccagcc</u>	<u>cgcttgctag</u>
13601	<u>aatctagagt</u>	<u>tgcctttgac</u>	<u>ccttggcccc</u>	<u>agccagcccc</u>	<u>gtgaccttgc</u>
13651	<u>ccgggagaag</u>	<u>gaggtggcct</u>	<u>ggagagctgc</u>	<u>tgtctccagc</u>	<u>cgccgcctgt</u>
13701	<u>ctccacagTA</u>	<u>TGGCAAGAAG</u>	<u>AAACGCAGAT</u>	<u>ACCTACCCTA</u>	<u>CAACCAGCAG</u>
13751	<u>CACCTGTACT</u>	<u>TCTTCTGAG</u>	<u>tgagtgtcca</u>	<u>tctgtccttc</u>	<u>tggggtgggg</u>
13801	<u>gagtgcctgg</u>	<u>gcctgcactg</u>	<u>tcctccctgc</u>	<u>tgtcctggac</u>	<u>cactcccagc</u>
13851	<u>cacttcctgg</u>	<u>ggcggggcac</u>	<u>gtctgtcagg</u>	<u>tctccctggg</u>	<u>catggcatcc</u>
13901	<u>tcccagcctc</u>	<u>tgcagtctgt</u>	<u>acacactctc</u>	<u>ccagcagcat</u>	<u>cgctttgccc</u>
13951	<u>cagctgtctc</u>	<u>ccgtgcctgg</u>	<u>gacaccttgc</u>	<u>agccacgggc</u>	<u>catcacagcc</u>
14001	<u>ctgctgggag</u>	<u>cttccccaa</u>	<u>ccccacgtag</u>	<u>aatttcttct</u>	<u>tgcctcact</u>
14051	<u>agagtggtec</u>	<u>ggagccctag</u>	<u>agtctttggg</u>	<u>cagttgttgg</u>	<u>ggcggacaga</u>
14101	<u>gtgaggactc</u>	<u>aagtctggcc</u>	<u>ctgacttgcg</u>	<u>gtgaagggtg</u>	<u>gtgggaggtg</u>
14151	<u>gtggggtaag</u>	<u>ggcagcctgg</u>	<u>ggaggcttgg</u>	<u>acacagaatt</u>	<u>gggggtgata</u>
14201	<u>tggggtcatt</u>	<u>cagctggatg</u>	<u>tgaccagcac</u>	<u>caacgtccca</u>	<u>ggggcattcc</u>
14251	<u>tggagtaaca</u>	<u>gagccccca</u>	<u>ctctggcgcc</u>	<u>cactcacctt</u>	<u>ggcagcccag</u>
14301	<u>ccccactcct</u>	<u>gaacactctc</u>	<u>atgccccctc</u>	<u>ttgcagTCGG</u>	<u>CCCGCCGCTG</u>
14351	<u>CTCACCTTGG</u>	<u>TGAACTTTGA</u>	<u>AGTGGAAAAT</u>	<u>CTGGCGTACA</u>	<u>TGCTGGTGTG</u>
14401	<u>CATGCAGTGG</u>	<u>GCGgtgagt</u>	<u>gggttgccca</u>	<u>ggaccccggg</u>	<u>catacggctg</u>
14451	<u>ccgtggcagg</u>	<u>aggtggtgcc</u>	<u>tcggggggaca</u>	<u>gtacctgccc</u>	<u>atgaaggcaa</u>
14501	<u>acagggtgca</u>	<u>catgtgcgtg</u>	<u>caacagtgtg</u>	<u>gctcacatgt</u>	<u>atgcggtgcaa</u>
14551	<u>cagtgtggct</u>	<u>cacatgtgtg</u>	<u>cgcgagcag</u>	<u>gagagcgagt</u>	<u>gtgcccggtga</u>
14601	<u>ctgtacgtgt</u>	<u>ggtggggggg</u>	<u>ggttgaggaa</u>	<u>cagggggggg</u>	<u>gtgggtctct</u>
14651	<u>ctcggtgagg</u>	<u>gtgtcttccc</u>	<u>aggaggagtt</u>	<u>gctggggcca</u>	<u>ctctgccagg</u>
14701	<u>catctgtgtc</u>	<u>cctggcaggg</u>	<u>tcttccccaa</u>	<u>cacaccctgc</u>	<u>atgacacctt</u>
14751	<u>cgtcactaaa</u>	<u>atcagcctcg</u>	<u>tgagctggca</u>	<u>gggcaaggac</u>	<u>cctgttcctt</u>
14801	<u>tactcagctg</u>	<u>agaaaaccag</u>	<u>agagggtggt</u>	<u>ggcctgtcct</u>	<u>gggctctgag</u>
14851	<u>gcaaatcagg</u>	<u>cagaagggtt</u>	<u>ggatgcctga</u>	<u>ggtcctcctc</u>	<u>ccaccaccca</u>
14901	<u>ggcctccaga</u>	<u>cctccgggca</u>	<u>cctggagacc</u>	<u>tctcggtatc</u>	<u>gcctctgccc</u>
14951	<u>tcctctgcag</u>	<u>GATTTGCTCT</u>	<u>GGGCCGCCAG</u>	<u>CTTCTATGCC</u>	<u>CGCTTCTTCT</u>
15001	<u>TATCCTACCT</u>	<u>CCCCTTCTAC</u>	<u>GGCGTCCCTG</u>	<u>GGGTGCTGCT</u>	<u>CTTCTTTGTT</u>
15051	<u>GCTGTCAGgt</u>	<u>atggcagggg</u>	<u>gtggcgaggt</u>	<u>cacacacagg</u>	<u>cgacaggtga</u>
15101	<u>ccccactgc</u>	<u>agccccccac</u>	<u>cagagcttcc</u>	<u>cttttcccgt</u>	<u>ctgcagaatg</u>
15151	<u>gggccagtgg</u>	<u>tactgcctcc</u>	<u>ctggcttgct</u>	<u>ggtggaatca</u>	<u>cataaacaca</u>
15201	<u>agcgtggcag</u>	<u>gagcccagg</u>	<u>tcggtggtt</u>	<u>tagggagcgt</u>	<u>ggcctggcct</u>
15251	<u>gtaagtggcc</u>	<u>cggtgggtgt</u>	<u>cggagctgct</u>	<u>ctggactcag</u>	<u>cctcacagt</u>
15301	<u>gacactgctc</u>	<u>cattcagatt</u>	<u>ctttaaacac</u>	<u>tggcaagggg</u>	<u>gcgatggcca</u>
15351	<u>caatcctatt</u>	<u>gtacagataa</u>	<u>ggaagtcaag</u>	<u>gccacttggg</u>	<u>gacagctgct</u>
15401	<u>ctccagcctc</u>	<u>cactcagggt</u>	<u>gcctaagtgg</u>	<u>tgagctggac</u>	<u>ctagggcagt</u>
15451	<u>gcccagacct</u>	<u>ccccacagGG</u>	<u>TCCTGGAAAG</u>	<u>CCACTGGTTC</u>	<u>GTGTGGATCA</u>
15501	<u>CACAGATGAA</u>	<u>CCACATCCCC</u>	<u>AAGGAGATCG</u>	<u>GCCACGAGAA</u>	<u>GCACCGGGAC</u>
15551	<u>TGGGTCAGCT</u>	<u>CTCAGgtggg</u>	<u>cagcagggtt</u>	<u>ggggcccatc</u>	<u>ctgggtgggg</u>
15601	<u>tgggggggtcc</u>	<u>cagctaggag</u>	<u>ccagatggca</u>	<u>aagcagggat</u>	<u>gaggccctga</u>
15651	<u>cggggctgcc</u>	<u>aggtggggga</u>	<u>tgggtgccgt</u>	<u>gggtcaggga</u>	<u>tctgcaacgg</u>
15701	<u>cctcctcaca</u>	<u>tgtgccccgc</u>	<u>cggcttccgg</u>	<u>cagCTGGCAG</u>	<u>CCACCTGCAG</u>
15751	<u>CGTGGAGCCC</u>	<u>TCACCTTTCA</u>	<u>CCAACTGGTT</u>	<u>CAGCGGGCAC</u>	<u>CTCAACTTCC</u>
15801	<u>AGATCGAGCA</u>	<u>CCAGtgagt</u>	<u>tgggtgctgg</u>	<u>gggccagtg</u>	<u>gaggtgggga</u>
15851	<u>gggggtcctg</u>	<u>ggaggggatc</u>	<u>ctgggagggg</u>	<u>accggtgggt</u>	<u>ggggcctctc</u>

FIG.2F

8/19

15901 tctggaatct cccacttcag gtgccagcat acgctcccca cccccagCCT
 15951 CTTCCCCAGG ATGCCGAGAC ACAACTACAG CCGGGTGGCC CCGCTGGTCA
 16001 AGTCGCTGTG TGCCAAGCAC GGCCTCAGCT ACGAAGTGAA GCCCTTCCTC
 16051 ACCGCGCTGG TGGACATCGT CAGgtgaggc tgcagcccg cccctctggt
 16101 ctggtggctt ccccagggcc tatgcctacc cttgtccagg tcagcctcat
 16151 gctgagcccc cagggtccct gaggctttct gtccacgtcc catgcccttc
 16201 ctcccttccc cagccttcac gcacacagtg agaatttctg gagcacctac
 16251 tgcagactca caaacagcag tgcctgcggt gaggaggtct atgcaaacct
 16301 acccccaaag gctgagggaa aaaagctaac agatccagtt tctcagaagg
 16351 aaacacttaa cagggactca taaacagaag ccatgtctca gggccgggtg
 16401 cggtggctca cgctgtaat tccagcactt ggggaggctg aggtgggagg
 16451 atcacttgag gtcaggagtt cgagaccagc ctggccaaca tgggtgaaacc
 16501 ccgtctctac taaaaaaaaa aaaaaaaaaa aaaaacaaac aaaaattagc
 16551 tgggtgtggt ggcagggtgcc cataatccca gctacttggg aggtgagggg
 16601 aggagaatca cttgaactcg caggggcaga ggttgacagt ggtgagatt
 16651 gtgcctttgc agtccagcct gggcaacaga gcaagactct ctcaaaaaca
 16701 aacaaaaaaa ccatgtctca ggcagccaag agttgggaca tcccctcaca
 16751 cgccctctag aaagaaccct ctatatagca agcttttagg gtgaacccca
 16801 tgcagggtgt tcttatgaac ctggtgacca ctggagggtta gataagcgct
 16851 tacaagagga gggtatctat gccatgagct tggcattcag ggtcaagcat
 16901 cggtcatcag acagttttgc ttgaagatgg cattgccctt gtagcaatgc
 16951 aggtctctaga gagcttctg cctcttggga gctgatgttc cttccagcaa
 17001 aggaaacagc aagcaattaa aataacaaat aagtacatta cagaagatgg
 17051 gcaaaagaac aatgaaaagc ccctcagggt ggggacaggg gaggggaggg
 17101 gggcgggccag gcaggggagg cagtttctaa ataggtggta ggtggggcag
 17151 tattgacagg ctgacgtgtg agcagggaca gggaggaggg gagaggtctc
 17201 gccacaggga catctggcaa agagcggtca ggcagagggc acttgaccct
 17251 gaatgccaaag ctcatggcat agatagccga ggcaggcatg caggcactca
 17301 gagaagggac acgcccggct tgcactcttg aaagctgccc ctactgggaa
 17351 tgactggcgg gcaggagtcg aagtggaaaa ggagagcaga ggacactgca
 17401 gccatccagg cgaggggtga tggggctcag cccttgtggt caccttggag
 17451 gtgggggaaca gaggccagat tccaggtctt atacctctgc gccttgtac
 17501 acgctgttcc ccttacttgg ttgcccttcc ttcctgtgct ggtgttcaga
 17551 tgcccacttc tccttcatga tctctcccag cctgatgtc tgagcccctg
 17601 ccatttggca cagcccttta gagcgcttgg cacagggctt ctagcagat
 17651 tgttgacatt tctggctcca ctgcccata tcaggcccaa gatcgggtgg
 17701 gcaggttcca cgtcctctct gtccttgggt tgcagcgccc agcaggaggc
 17751 agcaatggag aactgggtgc aggagggaca ggcccaccca ggctcatgcc
 17801 tggacttggc cttggctgcc ctccagctcc cctacccgac acccgtcacc
 17851 ccggtctaga ttccattcca gagaatgagc attcagctgt tctcccaacc
 17901 caccctccag cccgcatcgc tgccctgccc cagggaaggg aaccacagg
 17951 gaatgggat ctccgctcac attaccatg ggggatacag ggtgttagg
 18001 atcttgcaac tgagctccta acacccaccc cactgccac cccacctcc
 18051 cagGTCCCTG AAGAAGTCTG GTGACATCTG GCTGGACGCC TACCTCCATC
 18101 AGTGAAGGCA ACACCCAGGC GGGCAGAGAA GGGCTCAGGG CACCAGCAAC
 18151 CAAGCCAGCC CCCGGCGGGA TCGATACCCC CACCCCTCCA CTGGCCAGCC
 18201 TGGGGGTGCC CTGCCTGCCC TCCTGGTACT GTTGCTCTCC CCTCGGCCCC
 18251 CTCACATGTG TATTCAGCAG CCCTATGGCC TTGGCTCTGG GCCTGATGGG
 18301 ACAGGGGTAG AGGGAAGGTG AGCATAGCAC ATTTTCCTAG AGCGAGAATT
 18351 GGGGGAAAGC TGTTATTTTT ATATTAAAT ACATTCAGAT GTATTATGGA
 18401 GT

FIG.2G

9/19

1	CTTCGCTTCCTCGGGGTCCTGCTCGGACCTCGGCCACCGCCTGGGATCC	50
51	CCAGGACTCGTGCGTGCAGCATGGGCGGCGTCGGGGAGCCGGGACCGCGG	100
1	M G G V G E P G P R	10
101	GAGGGACCCGCGCAGCCGGGGGCACCGCTGCCCACCTTCTGCTGGGAGCA	150
11	E G P A Q P G A P L P T F C W E Q	27
151	GATCCGCGCGCACGACCAGCCCCGGCGACAAGTGGCTGGTTCATCGAGCGCC	200
28	I R A H D Q P G D K W L V I E R R	44
201	GCGTCTACGACATCAGCCGCTGGGCACAGCGGCACCCAGGGGGCAGCCGC	250
45	V Y D I S R W A Q R H P G G S R	60
251	CTCATCGGCCACCACGGCGCTGAGGACGCCACGGATGCCTTCCGTGCCTT	300
61	L I G H H G A E D A T D A F R A F	77
301	CCATCAAGATCTCAATTTTGTGCGCAAGTTCCTACAGCCCCTGTTGATTG	350
78	H Q D L N F V R K F L Q P L L I G	94
351	GAGAGCTGGCTCCGGAAGAACCCAGCCAGGATGGACCCCTGAATGCGCAG	400
95	E L A P E E P S Q D G P L N A Q	110
401	CTGGTCGAGGACTTCCGAGCCCTGCACCAGGCAGCCGAGGACATGAAGCT	450
111	L V E D F R A L H Q A A E D M K L	127
451	GTTTGATGCCAGTCCCACCTTCTTTGCTTTCCTACTGGGCCACATCCTGG	500
128	F D A S P T F F A F L L G H I L A	144
501	CCATGGAGGTGCTGGCCTGGCTCCTTATCTACCTCCTGGGTCCTGGCTGG	550
145	M E V L A W L L I Y L L G P G W	160
551	GTGCCCAGTGCCCTGGCCGCCTTCATCCTGGCCATCTCTCAGGCTCAGTC	600
161	V P S A L A A F I L A I S Q A Q S	177
601	CTGGTGTCTGCAGCATGACCTGGGCCATGCCTCCATCTTCAAGAAGTCCT	650
178	W C L Q H D L G H A S I F K K S W	194
651	GGTGAACACGTGGCCCCAGAAGTTCGTGATGGGGCAGCTAAAGGGCTTC	700
195	W N H V A Q K F V M G Q L K G F	210

FIG.3A

10/19

701	TCCGCCCCACTGGTGGAACCTCCGCCACTTCCAGCACCACGCCAAGCCCCAA	750
211	S A H W W N F R H F Q H H A K P N	227
751	CATCTTCCACAAAGACCCAGACGTGACGGTGGCGCCCGTCTTCCTCCTGG	800
228	I F H K D P D V T V A P V F L L G	244
801	GGGAGTCATCCGTCGAGTATGGCAAGAAGAAACGCAGATACCTACCCTAC	850
245	E S S V E Y G K K K R R Y L P Y	260
851	AACCAGCAGCACCTGTACTTCTTCCTGATCGGCCCGCCGCTGCTCACCT	900
261	N Q Q H L Y F F L I G P P L L T L	277
901	GGTGAACCTTTGAAGTGGAATCTGGCGTACATGCTGGTGTGCATGCAGT	950
278	V N F E V E N L A Y M L V C M Q W	294
951	GGGCGGATTTGCTCTGGGCCGCCAGCTTCTATGCCCGCTTCTTCTATCC	1000
295	A D L L W A A S F Y A R F F L S	310
1001	TACCTCCCCTTCTACGGCGTCCCTGGGGTGCTGCTCTTCTTTGTGCTGT	1050
311	Y L P F Y G V P G V L L F F V A V	327
1051	CAGGGTCCTGGAAAGCCACTGGTTTCGTGTGGATCACACAGATGAACCACA	1100
328	R V L E S H W F V W I T Q M N H I	344
1101	TCCCCAAGGAGATCGGCCACGAGAAGCACCGGGACTGGGTCAGCTCTCAG	1150
345	P K E I G H E K H R D W V S S Q	360
1151	CTGGCAGCCACCTGCAACGTGGAGCCCTCACTTTTCACCAACTGGTTCAG	1200
361	L A A T C N V E P S L F T N W F S	377
1201	CGGGCACCTCAACTTCCAGATCGAGCACCACCTCTTCCCCAGGATGCCGA	1250
378	G H L N F Q I E H H L F P R M P R	394
1251	GACACAACCTACAGCCGGGTGGCCCCGCTGGTCAAGTCGCTGTGTGCCAAG	1300
395	H N Y S R V A P L V K S L C A K	410
1301	CACGGCCTCAGCTACGAAGTGAAGCCCTTCTCACC GCGCTGGTGGACAT	1350
411	H G L S Y E V K P F L T A L V D I	427
1351	CGTCAGGTCCCTGAAGAAGTCTGGTGACATCTGGCTGGACGCCTACCTCC	1400
428	V R S L K K S G D I W L D A Y L H	444

FIG.3B

11/19

1401	ATCAGTGAAGGCAACACCCAGGCGGGCAGAGAAGGGCTCAGGGCACCAGC	1450
445	Q	445
1451	AACCAAGCCAGCCCCCGGCGGGATCGATACCCCACCCCTCCACTGGCCA	1500
1501	GCCTGGGGGTGCACTGCCTGCCCTCCTGGTACTGTTGTCCTCCCCCTCGGC	1550
1551	CCCCTCACATGTGTATTTCAGCAGCCCTATGGCCTTGGCTCTGGGCCTGAT	1600
1601	GGGACAGGGGTAGAGGGAAGGTGAGCATAGCACATTTTCCTAGAGCGAGA	1650
1651	ATTGGGGGAAAGCTGTTATTTTTATATTAATAACATTCAGATGTAAAAA	1700

FIG.3C

12/19

1	GTACAGCGGCAATGGGCGGTGTCGGGGAGCCCGGAGGGGGACTCGGGCCG	50
1	M G G V G E P G G G L G P	13
51	CGGGAGGGGCCCCGCACCGCTGGGGGCGCCCCTACCCATCTTCCGCTGGGA	100
14	R E G P A P L G A P L P I F R W E	30
101	GCAGATCCGCCAGCATGACCTACCAGGCGACAAGTGGCTGGTCATCGAGC	150
31	Q I R Q H D L P G D K W L V I E R	47
151	GCCGTGTCTACGACATCAGCCGCTGGGCACAGCGGCACCCAGGGGGTAGC	200
48	R V Y D I S R W A Q R H P G G S	63
201	CGCATCATCGGCCACCACGG	220
64	R I I G H H	69

FIG.4

13/19

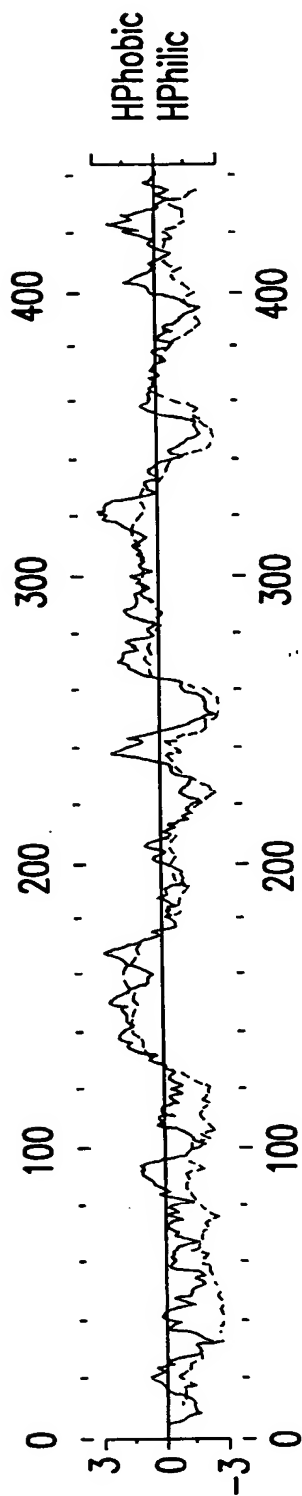


FIG. 5A

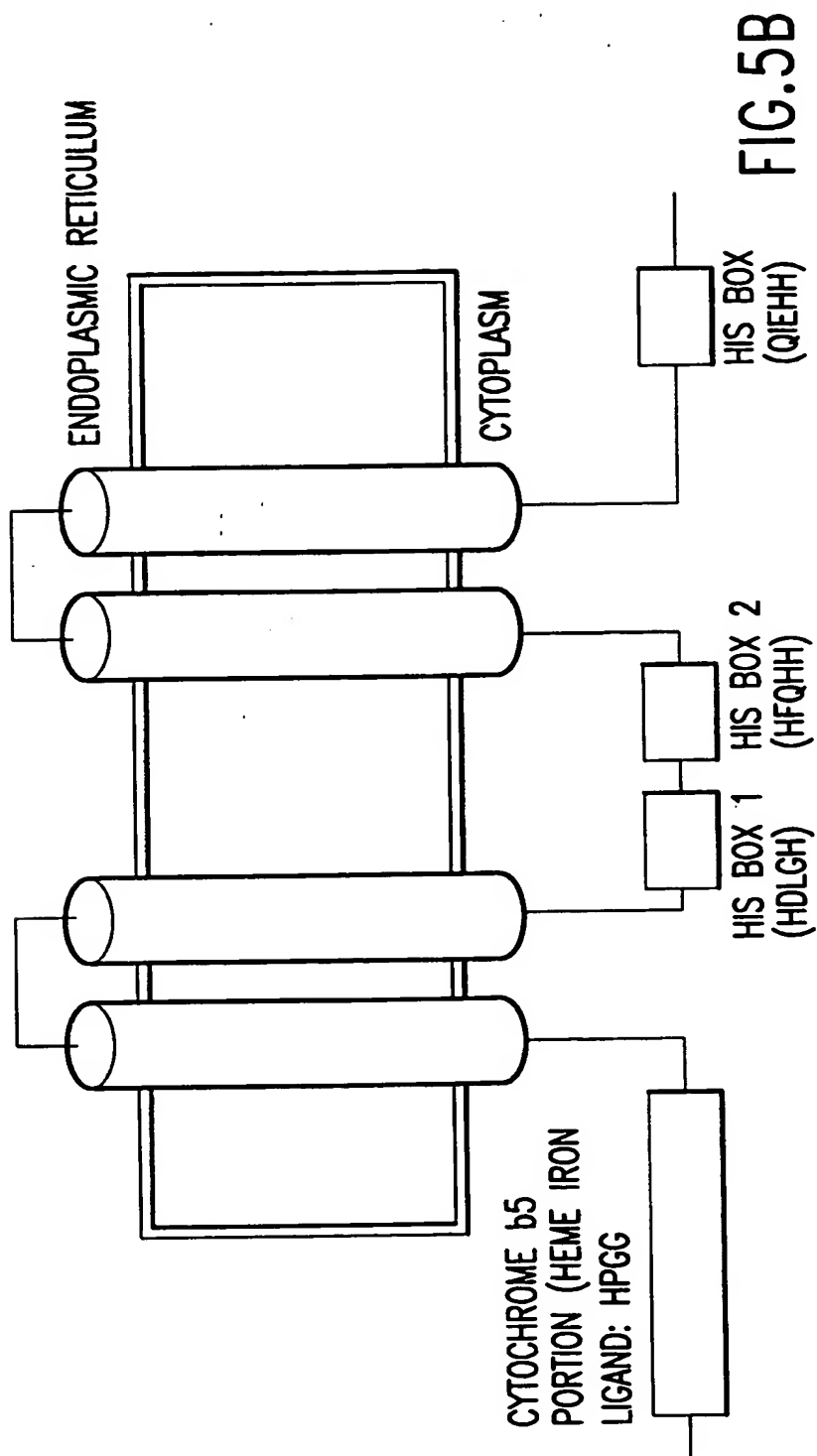


FIG. 5B

14/19

PROFILESCAN of : CYB5rp_correct_protein check: 5714 from: 1 to: 445

GETSEQ from bmd, December 2, 1997 14:20.

Compare to profile library: GenRunData:profilesca.n.fil

Profile: profiledir:cytochrome_b5.prf

Gap weight: 4.50 Gap Length weight: 0.05

Ave match: 0.27 Ave mismatch : -0.21

(Peptide) PROFILEMAKE v4.40 of: 0191.Msf2{*} Length: 48

Sequences: 24 MaxScore: 27.58 December 2, 1992 00:07

This profile is derived from PROSITE release 10.0 and has been tested by a database search against SWISS-PROT release 26.0. A comparison of the SWISS-PROT annotation and the results of the database search follows. For further information about this motif, consult the . . .

Profile: profiledir:cytochrome_b5.prf alignment: 1

Quality: 20.77 Gaps: 0

Ratio: 0.43 Length: 48

Normalized quality: 2.91

```


S      31 HDQPGDKWLVIERRVYDISRWAQRHPGGSRLIGHGAEDATDAFRAFH 78
      |: .. ||||. .|||::| . ||||. | .||.:||.| ::|
P      1 HNDGEETWLVNGQVYDITKFLLEHPGGPDVIMEAAGTDATEEFEAIH 48
  
```

```

*****
*Cytochrome b5 family, heme-binding domain signature *
*****
  
```

FIG.6

15/19

 pir:s68358 hypothetical protein - common sunflower
Length = 458

Score = 169 (79.4 bits), Expect = $2.8e-42$, Sum P(4) = $2.8e-42$
Identities = 31/85 (36%), Positives = 49/85 (57%) His box 3

Query: 348 IGHEKHRDWSSQLAATCNVEPSLFTNWFSGHLNFQIEHHLFPRMPRHNYSRVAPLVKSL 407
+G K +W Q T ++ S + +WF G L FQ+EHHLFPR+PR + ++P+ + L
Sbjct: 348 VGPPKGDWFEKQTRGTIDACSSWMDWFFGGLQFQLEHHLFPRLPCHLRSISPICREL 407

Query: 408 CAKHGLSYEVKPFLTALVDIVRSK 432
C K+ L Y F A V ++L+
Sbjct: 408 CKKYNLPYVLSFYDANVTTLKTLR 432

Score = 133 (62.5 bits), Expect = $2.8e-42$, Sum P(4) = $2.8e-42$
Identities = 21/53 (39%), Positives = 35/53 (66%)

HPGG motif
Query: 26 EQIRAHDPQDKWLVIERRVYDISRWAQRHPGGSRLLGHGAEDATDAFRAFH 78
++++ H+ P D W+ I +VY+++ WA+ HPGG + + +D TDAF AFH
Sbjct: 22 KELKKHNPNDLWISILGKVYNVTEWAKEHPGGDAPLINLAGQDVTDAFIAFH 74

Score = 118 (55.5 bits), Expect = $2.8e-42$, Sum P(4) = $2.8e-42$
Identities = 25/76 (32%), Positives = 34/76 (44%)

His box 1 His box 2
Query: 165 LAAFILAIISQAQSWCLOHDLGHASIFKKSWWNHVAQKFVMGQLKGFSAHWWNFRRHFQHEA 224
L+ IL ++ Q L HD GH + WN A F+ + G S WW + H HH
Sbjct: 152 LSGAILGLAWMQIAYLGHDAAGHYQMMATRGWNKFAGIFIGNCITGISIAWWKWTHNAHI 211

Query: 225 KPNIFHKDPDVTVAPV 240
N DPD+ P+
Sbjct: 212 ACNSLDYDPDLQHLPM 227

Score = 34 (16.0 bits), Expect = $2.8e-42$, Sum P(4) = $2.8e-42$
Identities = 7/14 (50%), Positives = 9/14 (64%)

FIG. 7A

16/19

⏏ gp:bou79010 1 PID:g2062403 Borago officinalis delta 6 desaturase mRNA,
complete cds. (gb:U79010) (NID:2062402)
Length = 448

Score = 179 (84.1 bits), Expect = 2.3e-42, Sum P(3) = 2.3e-42
Identities = 34/87 (39%), Positives = 48/87 (55%)

His box 3

Query: 348 IGHEKHRDWSSQLAATCNVEPSLFTNWFSGHLNFQIEHHLFPRMPRHNYSRVAPLVKSL 407
+G K +W Q T ++ + +WF G L FQIEHHLFP+MPR N +++P V L
Sbjct: 338 VGKPKGNNWFQKQTDGTLDISCPPWMDWFHGGLOFQIEHHLFPMPCNLRKISPYVIEL 397

Query: 408 CAKHGLSYEVKPFALTALVDIVRSLKKS 434
C K H L Y F A +R+L+ +
Sbjct: 398 CKKHNLPPYASFSKANEMTLRLTLRNT 424

Score = 144 (67.7 bits), Expect = 2.3e-42, Sum P(3) = 2.3e-42
Identities = 23/53 (43%), Positives = 36/53 (67%)

HPGG MOTIF

Query: 26 EQIRAHDPGDKWLVIERRVYDISRWAQRHPGGSRLIGHGAEDATDAFRAFH 78
++++ HD+PGD W+ I+ + YD+S W + HPGGS + ++ TDAF AFH
Sbjct: 12 DELKNHDKPGDLWISIQGKAYDVSDWVKDHPGGSFPLKSLAQEVTDAFVAFH 64

Score = 105 (49.3 bits), Expect = 2.3e-42, Sum P(3) = 2.3e-42
Identities = 22/68 (32%), Positives = 28/68 (41%)

His box 1


His box 2

Query: 176 QSWCLOHDLGHASIFKKSWWNHVAQKFMGQLKGFSAHWWNFRHFOHHAKPNIFHKDPDV 235
QS + HD GH + S N F L G S WW + H HH N DPD+
Sbjct: 153 QSGWIGHDAGHYMVVSDSRLNKFMGIFAAANCLSGISIGWKKWNHNAHHIACNSLEYDPDL 212

Query: 236 TVAPVFLL 243
p ++
Sbjct: 213 QVIPFLVV 220

FIG. 7B

17/19

 pir:s35157 Delta(6)-desaturase - *Synechocystis* sp.
Length = 359

Score = 126 (59.2 bits), Expect = 9.0e-09, Sum P(2) = 9.0e-09
Identities = 21/54 (38%), Positives = 33/54 (61%)

His box 3

Query: 372 FTNWFSGHLNFQIEHILFPRM⁺PRHNYSRVAPLVKSLCAKHGLSYEVKPFLTALV 425
F NMF G LN Q+ HILFP + +Y ++ ++K +C + G+ Y+V P A +
Sbjct: 292 FWNWFCGGLNHQVTHILFPNICHIHYPQLENI IKDVCQEFQVEYKVYPTFKAAI 345

Score = 36 (16.9 bits), Expect = 9.0e-09, Sum P(2) = 9.0e-09
Identities = 6/15 (40%), Positives = 8/15 (53%)

His box 2

Query: 209 GFSAHWWNFRHFQHH 223
G S+ W +RH H
Sbjct: 113 GLSSFLWRYRHNYLH 127

FIG.8

18/19

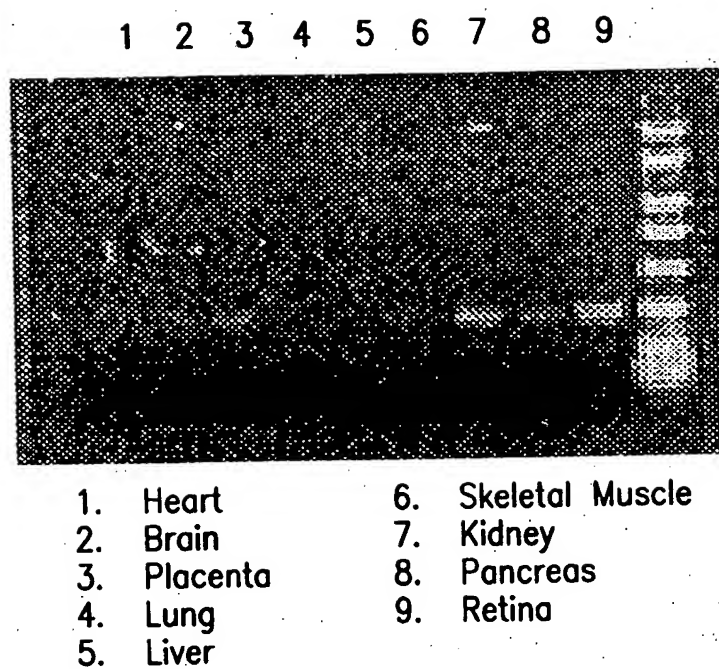


FIG.9A

19/19

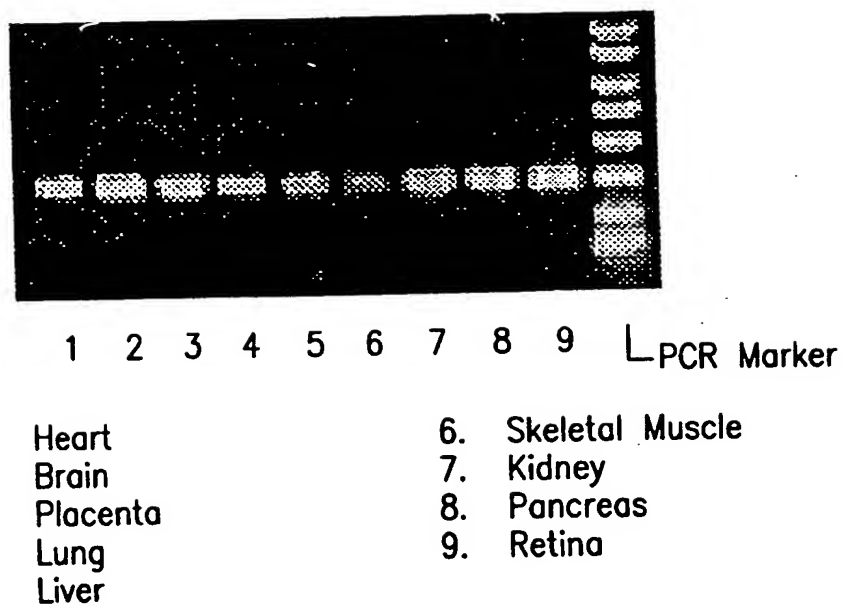


FIG.9B

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/23253

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : A61K 9/395; C12P 7/62; C12N 9/02, 15/00; C07H 19/00

US CL : 435/135, 189, 320.1, 452.3; 424/130.1; 536/23.2

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/135, 189, 320.1, 452.3; 424/130.1; 536/23.2

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
Please See Extra Sheet.

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Medline

Search terms: CYB5RP, delta-5 fatty acid desaturase, human or homo sapiens.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	Database GenBank, Accession AAC23396, submitted by LAMERDIN, JE, publicly available on 12 June 1998, see entire record.	1-15
X	Database GenBank, Accession AC004770, submitted by LAMERDIN, JE, publicly available on 12 June 1998, see entire record, especially identification of CDS at about line 50.	1-15
X,P	Database GenBank, Accession AAD31282 submitted by LI et al, publicly available on 19 May 1999, see entire record.	1-15
X	WO 98/39446 A2 (HUMAN GENOME SCIENCES, INC.) 11 September 1998, see entire document, especially SEQ ID No:63.	1-15

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:	*T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
A document defining the general state of the art which is not considered to be of particular relevance	*X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
B earlier document published on or after the international filing date	*Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
L document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	*Z* document member of the same patent family
O document referring to an oral disclosure, use, exhibition or other means	
P document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search

24 FEBRUARY 2000

Date of mailing of the international search report

15 MAR 2000

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

BRADLEY S. MAYHEW

Telephone No. (703) 308-0196

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/23253

B. FIELDS SEARCHED

Documentation other than minimum documentation that are included in the fields searched:

Because a CRF was not made available at the time of the search, Database GenBank Accession AF134404, which appears to encode the same desaturase as set forth in Figures 3A-C of the instant application, was searched against all available amino acid and nucleic acid databases.

This Page is inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☒ BLACK BORDERS
- ☒ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☒ FADED TEXT OR DRAWING
- ☒ BLURED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLORED OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REPERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images
problems checked, please do not report the
problems to the IFW Image Problem Mailbox**

THIS PAGE BLANK (USPTO)